

**University of Warsaw**  
Faculty of Mathematics, Informatics and Mechanics

**Tomasz Grześkiewicz**

Student no. 394317

**Mateusz Kobak**

Student no. 385760

**Iwona Kotlarska**

Student no. 394380

**Krzysztof Piesiewicz**

Student no. 385996

# Dataloading optimisation for deep learning on NVIDIA GPUs

**Bachelor's thesis  
in COMPUTER SCIENCE**

Supervisor:

**dr Janusz Jabłonowski**

University of Warsaw

Faculty of Mathematics, Informatics and Mechanics

Institute of Informatics

May 2020

## **Supervisor's statement**

Hereby I confirm that the presented thesis was prepared under my supervision and that it fulfils the requirements for the degree of Bachelor of Computer Science.

Date

Supervisor's signature

## **Authors' statements**

Hereby I declare that the presented thesis was prepared by me and none of its contents was obtained by means that are against the law.

The thesis has never before been a subject of any procedure of obtaining an academic degree.

Moreover, I declare that the present version of the thesis is identical to the attached electronic version.

Date

Authors' signatures

## **Abstract**

In this thesis performance of data loading process for deep learning on Nvidia GPUs has been analyzed. The overview of the currently used techniques of optimization are included. Over the course of the work, data loading process has been profiled to identify bottlenecks in the most common use cases in PyTorch and TensorFlow frameworks and chosen ones were mitigated, on either internal level of those frameworks or by creating examples of usage that improve the performance.

## **Keywords**

deep learning, GPU, dataloader

## **Thesis domain (Socrates-Erasmus subject area codes)**

11.3 Informatics, Computer Science

## **Subject classification**

D. Software

## **Tytuł pracy w języku polskim**

Optymalizacja wprowadzania danych na karty graficzne NVIDIA



# Contents

**Introduction** . . . . . 5



# Introduction

## Topic

For several years the terms *Deep learning* and *Neural networks* have been getting more and more recognisable thanks to the great impact of this technology on a huge variety of business aspects. *Speech recognition*, *interpretation of natural language*, *image classification*, *computer vision* and even *self-driving vehicles* are examples of the use of the deep learning technology. The increasing demand for deep learning solutions has caused the race to maximised the performance of available computing machines. The research and practise have shown that GPUs are best suited for deep learning computation purpose because they are unrivaled in linear algebra computations. There are several manners of increasing the performance and all of them are the interests of research and development teams of information technology industries. Some of the ways are constructing more powerful computing clusters or faster computation units. Others are targeted at software issues (operational systems, programming and algorithmic techniques). *Dataloading* mostly relates to the category which has been mentioned as the last one. Let assume the following definition.

*Dataloading* is the process of copying data from non-volatile storage to GPU memory including processing the data in CPU.

## Functionality