

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к дипломной работе

**«Реализация эффективного использования нейронных сетей на
графических процессорах»**

Бинцаровский Леонид Петрович

Научный руководитель – старший преподаватель кафедры дискретной
математики и алгоритмики ФПМИ Пирштук Д. И.

Минск, 2025

АННОТАЦИЯ

Дипломная работа, 49 страниц, 3 таблицы, 9 иллюстраций, 13 источников.

Ключевые слова: ИНФЕРЕНС НА GPU, C++, ГРАФИЧЕСКИЕ ПРОЦЕССОРЫ, GPU, OPENGL, VULKAN, ФРАГМЕНТНЫЙ ШЕЙДЕР, ВЕРШИННЫЙ ШЕЙДЕР, НЕЙРОННЫЕ СЕТИ, ESPCN, MEDIAPIPE, TFLITE.

Объект исследования является задача эффективного использования нейронных сетей на графических процессорах.

Предметом исследования являются методы и подходы для эффективного использования нейронных сетей на графических процессорах.

Целью работы является анализ задачи эффективного использования нейронных сетей на графических процессорах, разработка пайплайна для решения поставленной задачи.

Методами исследования являются системный подход, расчет количественных и качественных показателей, изучение соответствующей литературы и электронных источников, постановка задачи и её решение.

Полученные результаты и их новизна: были рассмотрены различные подходы к решению поставленной задачи и выявлены их основные недостатки. Успешно разработан и реализован пайплайн для поставленной задачи. Произведено сравнительное тестирование разработанного пайплайна с лидирующими в рамках поставленной задачи фреймворками.

Достоверность материалов и результатов дипломной работы: использованные материалы и результаты дипломной работы являются достоверными. Работа выполнена самостоятельно.

Областью возможного практического применения является автомобильная промышленность, медицина, бизнес, сфера безопасности, сфера развлечений.

АНАТАЦЫЯ

Дыпломная праца, 49 старонак, 3 табліцы, 9 ілюстрацый, 13 крыніцы.

Ключавыя словы: ИНФЕРЭНС НА GPU, C++, ГРАФІЧНЫЯ ПРАЦЭСАРЫ, GPU, OPENGL, VULKAN, ФРАГМЕНТНЫ ШЭЙДАР, ВЯРШЫННЫ ШЭЙДАР, НЕЙРОНАВЫЯ СЕТКІ, ESPCN, MEDIAPIPE, TFLITE.

Аб'ектам даследавання з'яўляецца задача эфектыўнага выкарыстання нейронавых сетак на графічных працэсарах.

Прадметам даследавання з'яўляюцца метады і падыходы фектыўнага выкарыстання нейронавых сетак на графічных працэсарах.

Мэтай даследвання з'яўляецца аналіз праблемы эфектыўнага выкарыстання нейрасетак на графічных працэсарах і распрацоўка пайплайна для яе рашэння.

Метадамі даследавання з'яўляюцца сістэмны падыход, разлік колькасных і якасных паказчыкаў, вывучэнне адпаведнай літаратуры і электронных крыніц, пастаноўка задачы і яе вырашэнне.

Атрыманыя вынікі і іх навізна: былі разгледжаны розныя падыходы да вырашэння задачы і выяўлены іх асноўныя недахопы. Паспяхова распрацаваны і рэалізаваны пайплайн для пастаўленай задачы. Праведзена параўнальнае тэсціраванне распрацаванага пайплайна з вядучымі фреймворкамі ў межах задачы.

Даставернасць матэрыялаў і вынікаў дыпломнай працы: выкарыстаныя матэрыялы і вынікі дыпломнай Працы з'яўляюцца даставернымі. Праца выканана самастойна.

Вобласцю магчымага практычнага прымянення з'яўляецца аўтамабільная прамысловасць, медыцына, бізнес, бяспека, сфера забаў.

ANNOTATION

Diploma work, 49 pages, 3 tables, 9 illustrations, 13 sources.

Keywords: INFERENCE ON GPU, C++, GRAPHICS PROCESSORS, GPU, OPENGL, VULKAN, FRAGMENT SHADER, VERTEX SHADER, NEURAL NETWORKS, ESPCN, MEDIAPIPE, TFLITE.

The object of the research the task of effective use of neural networks on graphic processors.

The subject of the research is methods and approaches for the effective use of neural networks on GPUs.

The purpose of the research is to analyse the problem of effective use of neural networks on graphic processors and to develop a pipeline for solving the task.

Methods of research are system approach, calculation of quantitative and qualitative indicators, study of relevant literature and electronic sources, problem statement and its solution.

The results of the work and their novelty: various approaches to solving the task were considered and their main shortcomings were identified. A pipeline for the task was successfully developed and implemented. Comparative testing of the developed pipeline against leading frameworks for the task was conducted.

Authenticity of the materials and results of the diploma work: the materials used and the results of the diploma work are authentic. The work has been put through independently.

Recommendations on the usage automotive industry, medicine, business, security, entertainment.