

## ИССЛЕДОВАНИЕ ВЛИЯНИЯ АЛГОРИТМА ВИНОГРАДА НА ТОЧНОСТЬ ОПЕРАЦИИ СВЕРТКИ

**Насыров К. Н., Перминов И. В.**

(Университет ИТМО, г. Санкт-Петербург)

**Научный руководитель – к.ф.-м.н., доцент Жданов Д. Д.**

(Университет ИТМО, г. Санкт-Петербург)

В данном исследовании рассматривается точность сверточных операций, основанных на алгоритме Винограда и использующихся в нейронных сетях, для чисел с плавающей точкой.

Сверточные вычисления являются одним из наиболее затратных элементов нейронных сетей и на протяжении многих лет публикуются исследования по увеличению их производительности для широкого спектра платформ.

Алгоритм Винограда является одним из наиболее эффективных способов ускорить вычисление сверток путем перевода вычислений во внутренний домен алгоритма, что позволяет сократить количество вычислений в несколько раз. Однако, в процессе перевода данных возможно искажение результатов свертки, связанное с умножением на дробные константы и округлениями результатов выполнения арифметических операций над числами с плавающей точкой. В рамках алгоритма процесс перехода в домен состоит из умножения на матрицы преобразования, содержание данных матриц не фиксировано и зависит от выбранных опорных точек, поэтому точность результата также зависит от выбранной матрицы преобразования.

В некоторых случаях нейронные сети могут быть крайне чувствительными к точности используемых операций и искажения результатов могут быть неприемлемыми, в связи с чем появилась необходимость оценки влияния алгоритма Винограда на результат свертки.

В ходе исследования используется несколько различных наборов матриц преобразования, происходит оценка влияния выбранных констант и возможной реализации алгоритма на точность сверточных вычислений.

Результаты работы могут быть использованы при выборе оптимального алгоритма свертки или улучшении точности существующих.