

FPGA реализация нейронной сети прямого распространения для распознавания рукописных чисел

Е.А. Кривальцевич М.И. Вашкевич

krivalcevi4.egor@gmail.com, vashkevich@bsuir.by

Белорусский государственный университет
информатики и радиоэлектроники
Кафедра электронных вычислительных средств



XIV Международная научная конференции
«Информационные технологии и системы»
Минск, Республика Беларусь



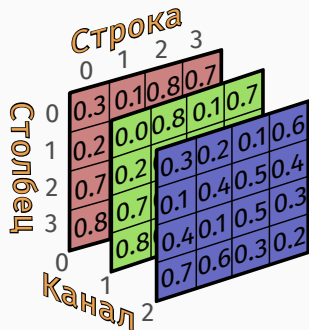
20 ноября, 2024

1. Прототипирование нейронных сетей на FPGA
2. Постановка задачи
3. Обучение нейронной сети
4. Аппаратная реализация нейронной сети
5. Использование PYNQ для прототипирования и тестирования нейронной сети
6. Описание эксперимента и результаты

Введение

Прототипирование нейронных сетей на FPGA

- В настоящее время наблюдается интерес к использованию кватернионов при построении нейронных сетей для обработки многомерных данных^a
- Цветные изображения являются важным примером многомерных данных
- Обычно RGB-изображения обрабатываются при помощи сверточных нейронных сетей. Входное изображение интерпретируется, как 3-х мерный тензор



^aKusakabe, T., Kouda, N., Isokawa, T., Matsui, N. : A Study of Neural Network Based on Quaternion. Proceeding of SICE Annual Conference (2002) 776–779

Цель исследования

- Получить базовую модель автокодировщика на основе кватернионов
- Выяснить дает ли преимущество использование кватернионов в задаче сжатия изображений при помощи автокодировщика
- Оценить преимущества кватернионной нейронной сети над вещественнозначной в задаче сжатия изображений (MSE, PSNR, SIMM)

Обучение нейронной сети

Архитектура НС

Аппаратная реализация нейронной сети

Аппаратная реализация нейронной сети

Использование PYNQ для прототипирования и тестирования нейронной сети

Автокодировщик на основе вещественнозначной НС

RAE – real autoencoder

Автокодировщик на основе кватернионной НС

QAE – quaternion autoencoder

Эксперимент и результаты

MSE

PSNR

SIMM

Результат работы автокодировщика RAE-2048

Сравнение полученных результатов

- Полученные автокодировщики на основе кватернионной НС сравнивались с другими вещественнозначным автокодировщиком, имеющим архитектуру «бутылочное горлышко»¹.
- Автокодировщик¹, использует НС с двумя слоями: первый сверточный, а второй полносвязный и обеспечивает степень сжатия 2:1, т.е. внутренний слой содержал в два раза меньше элементов, чем входной и выходной.
- У автокодировщика¹ индекс структурной схожести (SIMM) равен 0,905.
- Полученный в данной работе автокодировщик QAE-1024 обеспечивает сжатие 3:1 и имеет SIMM равный 0,922.

¹Yijing Watkins и др. “Image compression: Sparse coding vs. bottleneck autoencoders”. В: 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI). IEEE. 2018, с. 17—20.

- Рассмотренный вычислительный эксперимент на основе полносвязного автокодировщика показывает, что представление скоррелированных данных, таких как цветные RGB-изображения, в алгебре кватернионов позволяет лучше учитывать характер исходных данных.
- Предложенные автокодировщики, имеющие различную степень сжатия, позволяют получить более высокие значения объективных характеристик декодирования цветных изображений по сравнению с аналогичными вещественнозначными автокодировщиками (PSNR в среднем выше на 3,85 дБ, SSIM на 0,18).