ИССЛЕДОВАНИЕ ЭФФЕКТОВ КВАНТОВАНИЯ ПАРАМЕТРОВ НЕЙРОННОЙ СЕТИ НА ОСНОВЕ ОБУЧАЕМОГО РАЗДЕЛИМОГО ПРЕОБРАЗОВАНИЯ

Кривальцевич Е.А. **Вашкевич М.И.**
krivalcevi4.egor@gmail.com, vashkevich@bsuir.by

Белорусский государственный университет информатики и радиоэлектроники
Минск, Беларусь

XV Международная научная конференции «Информационные технологии и системы»

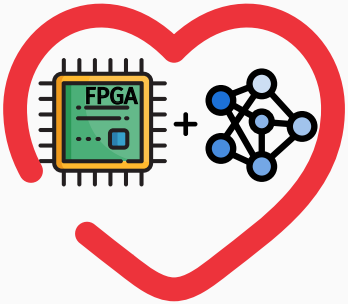
Содержание доклада

1. Задача реализации нейронных сетей (НС) на FPGA
2. Особенности существующих архитектур НС для FPGA
3. Разделимое преобразование
4. Обучаемое двумерное разделимое преобразование (ОДРП)
5. FPGA-реализация НС на основе ОДРП
6. Результаты экспериментов
7. Выводы

Введение

Реализация нейронных сетей (НС) на FPGA

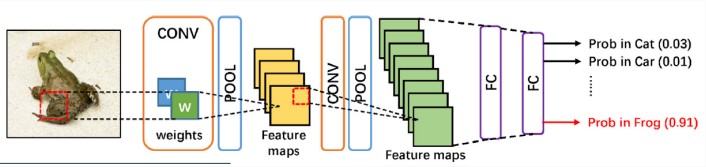
- Ускорение процесса исполнения (англ. *inference*) НС.
- Прототипирование и верификация аппаратной архитектуры НС перед реализацией в ASIC
- Поиск баланса между точностью работы НС и потреблением аппаратных ресурсов (оптимальное квантование коэффициентов)
- Разработка архитектур НС, ориентированных на FPGA/ASIC



Особенности архитектур существующих НС для FPGA

Особенности НС с точки зрения реализации на FPGA

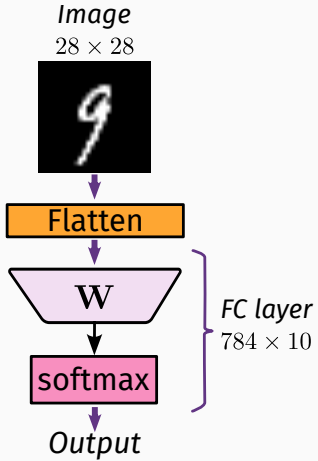
- **Многослойные перцептроны** – имеют простую, регулярную структуру, которая хорошо отображается на архитектуру FPGA.
  Недостатки: невысокая производительность, большое число параметров модели, нагрузка на память.
- **Сверточные НС** – имеют более высокую производительность, меньшее число параметров, допускают параллельную обработку.
  Недостатки: сложность управления, веса представляют собой тензоры и имеют сложный паттерн доступа к памяти.
- Чаще всего используются гибридные архитектуры.[1]



[1]S. Liang et al. “FP-BNN: Binarized neural network on FPGA”. In: *Neurocomputing* 275 (2018), pp. 1072–1086.

Пример НС для распознавания рукописных цифр

- Задача: реализации на FPGA НС для распознавания рукописных цифр (база MNIST)
- Простой однослойный перцептрон (7,8 тыс. параметров) позволяет достичь относительно невысокой точности 92,5%[a]
- При добавлении одного скрытого слоя на 40 нейронов число параметров НС увеличивается до 31,8 тыс.



[a] Е. Кривальцевич и М. Вашкевич. “Исследование аппаратной реализации нейронной сети прямого распространения для распознавания рукописных цифр на базе FPGA”. В: 23.2 (2025), с. 101—108.

Разделимое преобразование

Двумерное разделимое преобразование

- **Двумерные разделимые преобразования** применяются в обработке изображений для снижения вычислительносй сложности при пространственной фильтрации. Ядро преобразования имеет вид:

$$\mathbf{W} = \mathbf{v} \times \mathbf{h}^T,$$

где $\mathbf{W} \in \mathbb{R}^{n \times n}$, $\mathbf{v}, \mathbf{h} \in \mathbb{R}^{n \times 1}$.

- Разделимое преобразование $\mathbf{W}$ имеет $2n$ независимых параметров, вместо n^2 параметров, которые имеет обычное преобразование.
- Пример разделимого преобразования – фильтр Собеля:

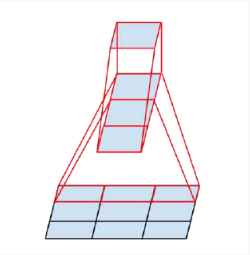
$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}.$$

Декомпозиция 2D-свертки

Идея использования декомпозиции сверток $n \times n$ на два последовательных слоя сверток $n \times 1$ и $\times n$ была предложена Сегеди[2] для совершенствования Inception-модуля НС GoogLeNet.

☞ Пример замены свертки 3×3 сверткой 3×1, которая дает 3 выхода, и сверткой 1×3, которая дает один выход.

- **Основная идея**: сокращение числа параметров и вычислительной сложности.



[2] C. Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

Обучаемое двумерное разделимое преобразование (ОДРП)

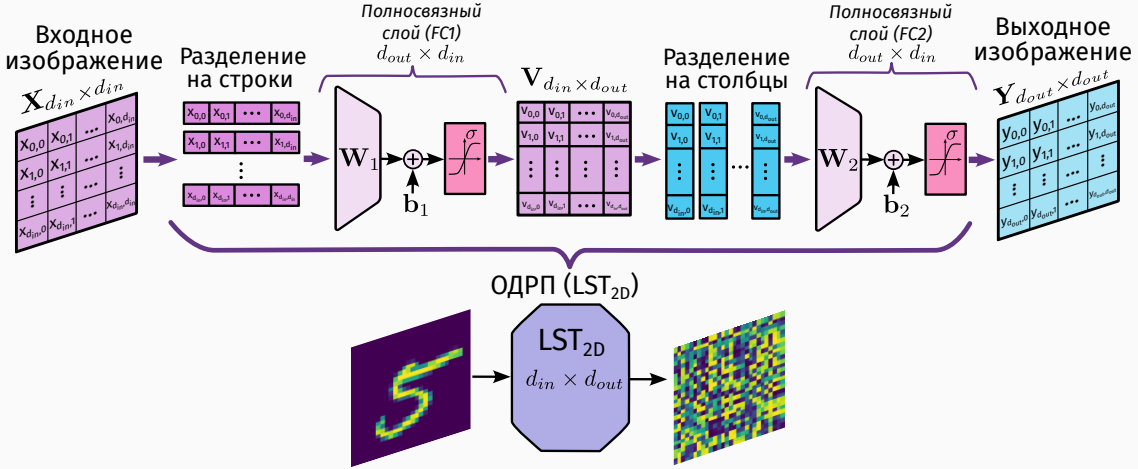
- Авторами[3] предложено обучаемое преобразование (LST), которое обрабатывает изображение вначале по строкам, а затем по столбцам, с использованием полносвязных слоев.
- Преобразование LST обрабатывает изображение $\mathbf{X}$ размера $d_{in} \times d_{in}$ при этом на выходе получается изображение $\mathbf{Y}$ размера $d_{out} \times d_{out}$:

$$\mathbf{Y} = \mathrm{LST}(\mathbf{X}) = \sigma(\mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{X}^T)^T),$$

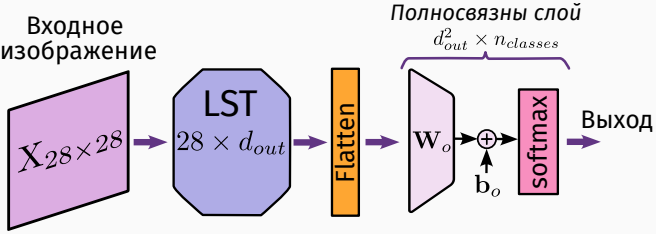
где $\mathbf{W}_1$, $\mathbf{W}_2$ – матрицы весов полносвязных слоев $\mathrm{FC1}$ и $\mathrm{FC2}$, соответственно, а d_{in} и d_{out} – это параметры преобразования, $\sigma()$ – активационная функция.

[3] M. Vashkevich and E. Krivalcevich. “Compact and Efficient Neural Networks for Image Recognition Based on Learned 2D Separable Transform”. In: *2025 27th International Conference on Digital Signal Processing and its Applications (DSPA)*. 2025, pp. 1–6.

Обучаемое двумерное разделимое преобразование (ОДРП)



Нейронная сеть LST-1



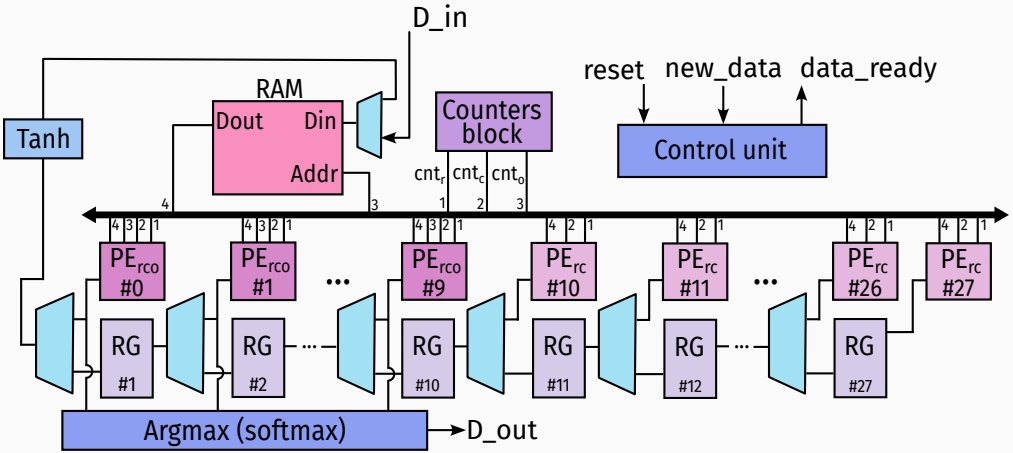
- LST можно рассматривать, как базовый блок для построения компактных нейронных сетей для распознавания изображений
- LST-1 – простейший вариант нейронной сети, использующий блок LST.
- Число параметров LST-1:

$$N_{params} = 2 \cdot (d_{in} + 1) \cdot d_{out} + (d_{out}^2 + 1) \times 10$$

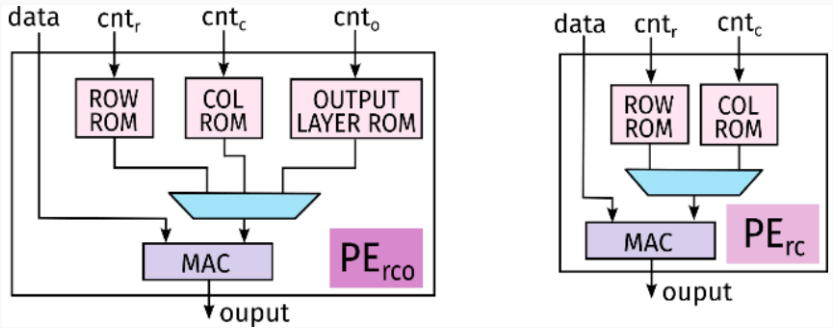
- Для $d_{in} = d_{out} = 28$ число параметров модели $N_{params} = 9\ 474$.

Реализация НС LST-1 на FPGA

Реализация LST-1 на FPGA



Процессорные элементы (ПЭ) LST-1



- Реализация LST-1 включает 10 ПЭ PE_{rco} и 18 ПЭ PE_{rc}.
- На первой стадии при вычислении LST используются все ПЭ. В каждом ПЭ хранится одна строка матрицы $\mathbf{W}_1$ в ПЗУ «ROW ROM», и одна строка матрицы $\mathbf{W}_2$ в ПЗУ «COL ROM».
- На второй стадии выполняется расчет выходов классифицирующего слоя, используются ПЭ типа rco, задействуется память «OUTPUT LAYER».

Аппаратные завтраты FPGA на реализацию LST-1

- IP-ядро HC LST-1 протестировано на отладочной плате Zybo Z7 (СнК XC7Z010) с использованием фреймфорка PYNQ.
- В таблице приведены затраты на IP-ядро LST-1, реализованное в 13-разрядной арифметике с фиксированной запятой.
- Тактовая частота IP-блока 48,8 МГц
- Число тактов на распознавание одного изображения: 4934 такта.
- Пропускная способность: 9890 изображений/с.

Тип ресурса	Использовано	Доступно	% Использования
LUT	914	17600	5,3
Flip Flop	302	35200	0,86
RAMB18	67	120	55,83
DSP	57	80	71,25

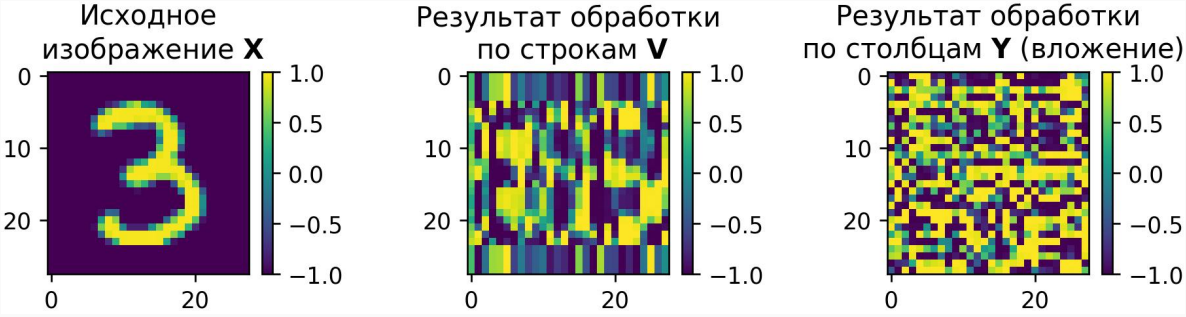
Эксперименты & результаты

Обучение НС LST-1

- Набор данных MNIST содержит (60k + 10k изображений рукописных цифр размера 28 × 28)
- Инициализация весов модели выполнялась с использованием метода Ксавье
- Целевая функция – отрицательный логарифм правдоподобия (`torch.nn.NLLLoss`)
- Обучение выполнялось при помощи алгоритма Adam (скорость обучения регулировалась при помощи планировщика по методу косинусного отжига, число эпох обучения – 300, размер батча – 1000)
- Для оценки качества распознавания использовался показатель точности (англ. *accuracy*)

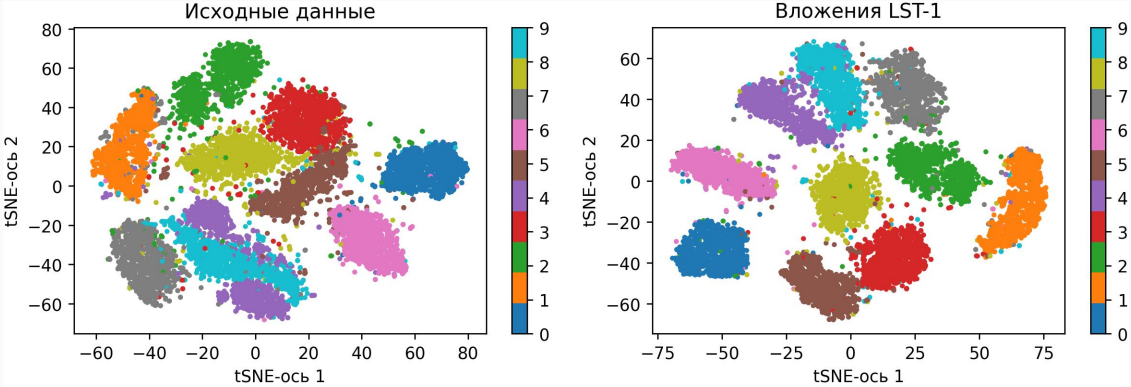
Вложения модели LST-1

- Модель LST-1 кодирует изображение в виде нерегулярного паттерна, похожего на QR-код.



Анализ вложений модели LST-1

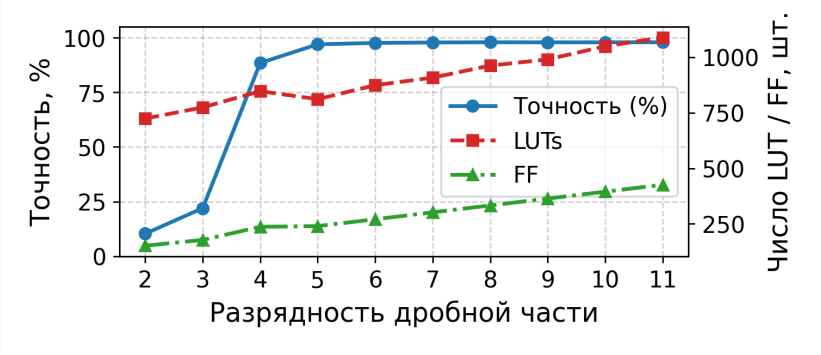
- Представления изображений, получаемые на скрытом слое модели LST-1, были проанализированы с использованием нелинейного метода понижения размерности tSNE.



Результаты экспериментов

- Матрицы спутанности, полученные с помощью модели с плавающей запятой и реализации на FPGA, согласуются результатами, полученными на Python-модели с фиксированной запятой.
- **Точность модели LST-1** с весами, квантованными в формате Q6.7 – **98,28%**.

Результаты экспериментов

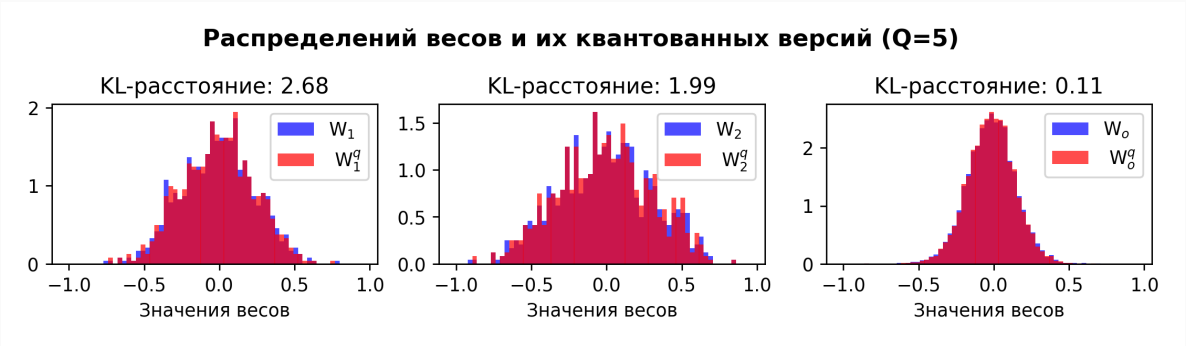


- Точность более 97% достигается начиная с 5 бит в дробной части.

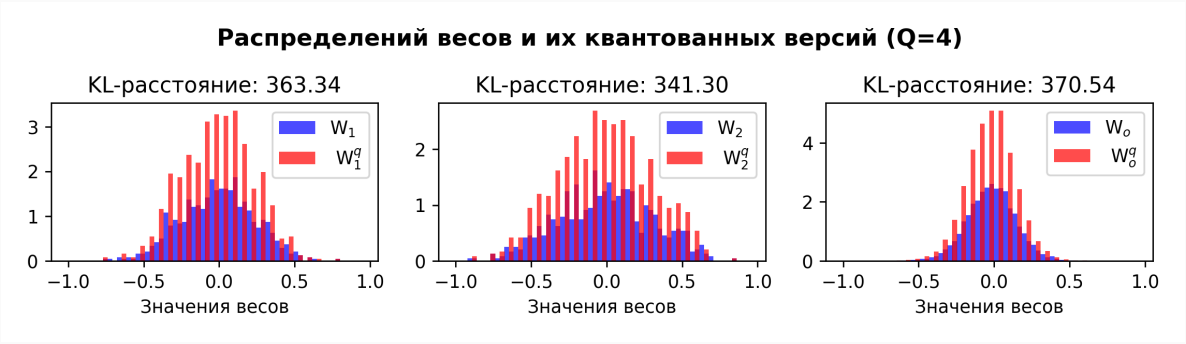
Анализ квантованных весов модели LST-1

Для анализа весов модели LST-1 до и после квантования использовалось расстояние Кульбаха-Лейблера:

$$\mathrm{KL}(p, q) = \sum_x p(x) \log \frac{p(x)}{q(x)}.$$



Анализ квантованных весов модели LST-1



- После перехода от 5 к 4 разрядам в дробной части KL-расстояние между распределениями резко увеличивается.
- Точность модели LST-1 при переходе от 5 к 4 разрядам в дробной части падает с 88,5% до 21,9%.

Сравнение архитектур НС для MNIST

Автор	Модель	Число параметов	Точность, %
Medus [4]	784-600-600-10	891 610	98,63%
Samragh [5]	784-512-512-10	670 208	98,40%
Huynh [6]	784-126-126-10	115 920	98,16%
Huynh	784-40-40-40-10	34 960	97,20%
Westby [7]	784-12-10	9 550	93,25%
предлагаемая	LST-1	9 474	98,28%

[4] L. D. Medus et al. “A novel systolic parallel hardware architecture for the FPGA acceleration of feedforward neural networks”. In: *IEEE Access* 7 (2019), pp. 76084–76103.

[5] M. Samragh et al. “Customizing neural networks for efficient FPGA implementation”. In: *2017 IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM)*. 2017, pp. 85–92.

[6] T. V. Huynh. “Deep neural network accelerator based on FPGA”. In: *2017 4th NAFOSTED Conf. on Information and Comp. Sc.* 2017, pp. 254–257.

[7] I. Westby et al. “FPGA acceleration on a multilayer perceptron neural network for digit recognition”. In: *The Journal of Supercomputing* 77.12 (2021), pp. 14356–14373.

Заключение

- Предложено двумерное обучаемое разделяемое преобразование, которое может быть использовано в качестве базового блока для построения компактных нейронных сетей для распознавания изображений.
- Разработана модель LST-1, которая в задаче распознавания рукописных цифр, обеспечивает точность 98,37%, при наличии всего 9,4 тыс. параметров.
- Предложена аппаратная архитектура для реализации модели LST-1 на основе FPGA.
- Показано, что имеет место связь между KL-растоянием распределений квантованных и неквантованных весов модели и производительностью модели.