

УДК 004.8

**РАЗРАБОТКА МЕТОДИКИ ПОСТРОЕНИЯ ДИНАМИЧЕСКИХ СИСТЕМ ДЛЯ  
ВНЕСЕРВЕРНОЙ ОБРАБОТКИ ДАННЫХ**

Лобачев И.М., Малерик Р.П., Филягин Д.Г.  
д.т.н., проф. Антощук С. Г.

Одесский Национальный Политехнический Университет, УКРАИНА

**АННОТАЦИЯ.** В данной работе рассматривается разработка методики построения динамических систем для внесерверной обработки сигналов и изображений с использованием нейронных сетей и специализированных аппаратных ускорителей.

Все чаще появляется потребность в использовании нейронных сетей на устройствах без доступа к интернету. Привычные облачные вычисления не гарантируют безопасность передачи данных на сервер и ставят под угрозу ее конфиденциальность. Более того, удобно использовать компактное автономное устройство для интеллектуальной обработки большого объема данных с возможностью динамического расширения системы. Например, обработка изображений требует большой вычислительной мощности, а обработка потока изображений еще больше. Расширение области классификации и увеличение сложности нейронной сети может приводить к тому, что существующая система перестанет справляться с поставленной задачей за установленное время. Таким образом, необходимо предусматривать возможности динамического увеличения вычислительной мощности.

Целью работы является уменьшение количества затрачиваемых ресурсов в динамических системах внесерверной обработки данных, использующих нейронные сети.

Обработка данных может проводиться различными способами. На сегодняшний день сложным и в то же время очень мощным инструментом являются нейронные сети. Однако их использование может привести к нехватке вычислительных ресурсов. В таких случаях целесообразно использование специализированных устройств для проведения вычислений. На рынке представлены различные готовые решения:

MobileyeEyeQ — специализированный процессор, ускоряющий обработку алгоритмов машинного зрения для использования в беспилотном автомобиле.

GoogleTensorProcessingUnit (Google TPU) — тензорный процессор, относящийся к классу нейронных процессоров, являющийся специализированной интегральной схемой разработанной корпорацией Google и предназначенной для использования с библиотекой машинного обучения TensorFlow.

IntelNervanaNeuralNetworkProcessor (NNP) — первый коммерчески доступный тензорный процессор, предназначенный для постройки сетей глубокого обучения

IBM TrueNorth — нейроморфный процессор, построенный по принципу взаимодействия нейронов, а не традиционной арифметики.

IntelMovidiusMyriad 2 — многоядерный ИИ-ускоритель основанный на VLIW-архитектуре (VeryLongInstructionWord), с дополненными узлами, предназначенными для обработки видео. Он является основой MovidiusNeuralComputeStick – устройства с подключением по USB, реализующее технологию глубокого обучения нейронных сетей[1].

Для разработки предоставляется SoftwareDevelopmentKit (SDK) с примерами использования. На данный момент поддерживаются такие фреймворки, как TensorFlow и Caffe на языках программирования C++ и Python.

Проанализировав существующие аналоги, для создания прототипа собственной системы было решено использовать RaspberryPi 3 и нейрочипыMovidiusNeuralComputeStick, поскольку такое решение обеспечит мобильность, высокую скорость вычислений и сравнительно невысокую стоимость.

Проверка работы такой архитектуры с несколькими нейрочипами требует конкретного примера. В качестве источника данных для обработки были выбраны изображения с дорожными знаками в видеопотоке и данные электроэнцефалограммы (ЭЭГ)(рис. 1).

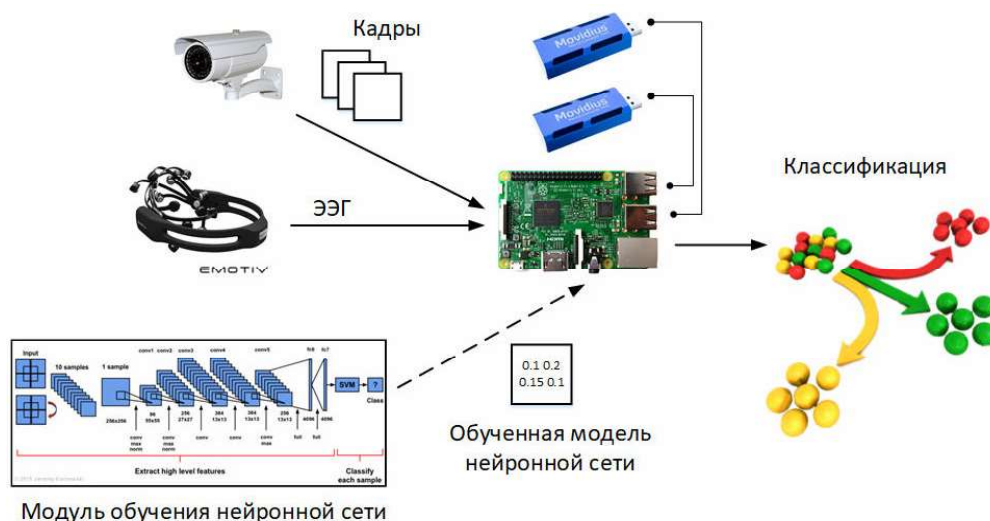


Рисунок 1 – Общая архитектура системы

В качестве тестовой выборки использованы изображения дорожных знаков из набора TheGermanTrafficSignRecognitionBenchmark, однако в будущем планируется расширить и дополнять эту выборку. Для распознавания изображений была использована классическая модель сверточной нейронной сети – AlexNet. Поскольку изображения, получаемые из видеопотока, подаются на вход сети без предварительного выделения на них дорожного знака, трудно оценить точность распознавания. Однако на данный момент изображения, на которых знак занимает 80-90% площади классифицируются с точностью 90-95%.

Для создания классификатора открытых/закрытых глаз человека был реализован перцептрон со структурой сети 14x40x1, который показал точность 94% на тестовой выборке[2], полученной с помощью нейрошлема Emotiv. Но так как для создания динамической системы было решено использовать нейрочипы Movidius NCS, а они работают только со сверточными нейронными сетями (CNN), датасет с сырыми данными ЭЭГ было решено преобразовать путем построения спектрограмм для перехода в частотную область путем преобразования Фурье и сохранить их в дальнейшем в виде изображений[3]. Полученные данные содержат диапазон частот от 0.5 до 30 Hz, так как именно этот диапазон отвечает за дельта (0.5 до 4 Hz), тэта (4 до 8 Hz), альфа (8 до 13 Hz) и бета (13 до 30 Hz) ритмы[4]. Такой подход может дать более высокую производительность, так как данные обрабатываются партиями, собранными за время 1-10 секунд. Хочется отметить, что показания по каждому из сенсоров предполагаются как независимые, что позволяет увеличить объем выборки для обучения.

### СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Intel выпускает MovidiusNeuralComputeStick — ускоритель ИИ в формате флэшки [Электронный ресурс]. – Режим доступа: URL: <https://www.ixbt.com/news/2017/07/21/intel-movidius-neural-compute-stick.html>
2. EEG EyeStateDataSet [Электронный ресурс]. – Режим доступа: URL: <http://archive.ics.uci.edu/ml/datasets/EEG+Eye+State>
3. Patients' EEG Data Analysis via Spectrogram Image with a Convolution Neural Network [Электронный ресурс]. – Режим доступа: URL: [http://www.springer.com/cda/content/document/cda\\_downloaddocument/9783319594200-c2.pdf?SGWID=0-0-45-1607908-p180877162](http://www.springer.com/cda/content/document/cda_downloaddocument/9783319594200-c2.pdf?SGWID=0-0-45-1607908-p180877162)
4. The Measurement of Brain Waves Network [Электронный ресурс].