

## **РАЗРАБОТКА СИСТЕМЫ ОБНАРУЖЕНИЯ И ОТСЛЕЖИВАНИЯ ОБЪЕКТОВ В ВИДЕОПОТОКЕ НА ОСНОВЕ ЛОКАЛЬНЫХ ОСОБЕННОСТЕЙ ИЗОБРАЖЕНИЯ ПРИ ОГРАНИЧЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСАХ**

**Фронин И. Р.<sup>1</sup>**

**Научный руководитель – Матяш А. В.<sup>2</sup>**

<sup>1</sup>Университет ИТМО

<sup>2</sup>ООО «Сэт Код»

nosowq10@gmail.com

### **Введение**

В общем случае модели обнаружения и отслеживания объектов можно разделить на классические и нейросетевые. Модели из первой группы, как правило, используются в устройствах с ограниченными ресурсами, так как не требуют сложных вычислительных операций и больших объёмов памяти. Однако классические модели существенно уступают нейросетевым в точности. Чтобы использовать модель нейронной сети без специализированного аппаратного ускорения, необходимо применять различные оптимизации. Актуальным и перспективным методом оптимизации является квантизация нейронных сетей [1] и, в частности, бинаризация [2]. Несмотря на популярность такого подхода, для задачи описания локальных областей изображения подобных решений недостаточно, что открывает возможности для разработки новых методов. При этом бинаризация и квантизация часто применяются к задаче классификации изображений, и исследования показывают, что данные методы способны снизить число производимых операций и размер модели в 64 раза при незначительной деградации метрик качества [3]. Поэтому данные подходы целесообразны для решения данной задачи и позволят значительно снизить аппаратные требования для устройств, которые могут использоваться, например, в составе беспилотных аппаратов и встраиваемых контроллеров.

### **Основная часть**

Решение данной задачи заключается в использовании модели бинарной нейронной сети для генерации бинарных векторов признаков (далее – дескрипторов) локальных областей изображения. Выбор локальных областей осуществляется с помощью классического алгоритма обнаружения угловых точек на изображении. На входной слой нейронной сети подаются яркости пикселей локальной области вокруг точки, а на выходе формируется N-битный дескриптор. Веса и активации нейронной сети бинаризуются с помощью функции знака и затем упаковываются в 64-битные числа по каналам. Свёртка осуществляется с помощью операций XNOR и подсчёта битов. Результат умножается на обучаемые параметры, соответствующие строкам, столбцам и каналам выходной активации. Первый и последний свёрточные слои архитектуры модели являются ключевыми, поэтому для сохранения точности не бинаризуются. Функция потерь должна обеспечивать максимальную информативность элементов дескриптора и их устойчивость к геометрическим и яркостным преобразованиям. Для этого оптимизируется баланс элементов дескриптора, снижается корреляция между битами внутри дескриптора, а также между дескрипторами различных областей в пределах батча, минимизируется расстояние Хэмминга между дескрипторами одной и той же преобразованной области изображения.

## Выводы

Предложенный подход был реализован и протестирован на данных датасета COCO2017 с применением аугментаций трёх уровней сложности, варьирующих угол поворота, масштаб, перспективные искажения и яркостные изменения. Проведено сравнение полноточной и бинарной версий модели. Для полноточной модели точность сопоставления составила 90% на простых примерах и 60% на сложных; для бинарной модели – 86% и 44% соответственно. Полученные результаты демонстрируют приемлемый компромисс между точностью и вычислительной эффективностью: снижение точности бинарной модели компенсируется многократным уменьшением объёма памяти и ускорением вычислений. Это делает предложенный подход перспективным для развёртывания на встраиваемых устройствах с ограниченными ресурсами, таких как микроконтроллеры, встраиваемые модули зрения и системы беспилотных аппаратов.

## Литература

1. Gholami A., Kim S., Dong Z., Yao Z., Mahoney M. W., Keutzer K. A Survey of Quantization Methods for Efficient Neural Network Inference // Low-Power Computer Vision: Improve the Efficiency of Artificial Intelligence. 2022. P. 291-326. <https://arxiv.org/abs/2103.13630>.
2. Haotong Qin, Ruihao Gong, Xianglong Liu, Xiao Bai, Jingkuan Song, Nicu Sebe. Binary Neural Networks: A Survey // Pattern Recognition. 2020. T. 105. № 7553.
3. Bulat A., Tzimiropoulos G. XNOR-Net++: Improved Binary Neural Networks // British Machine Vision Conference (BMVC). 2019. <https://arxiv.org/abs/1904.05858>.