УДК 004.93
DOI: 10.15593/2224-9397/2021.2.01

А.И. Тур, А.Н. Кокоулин

Пермский национальный исследовательский политехнический университет,
Пермь, Россия

ВОПРОСЫ ОПТИМИЗАЦИИ РАБОТЫ ПРОЕКТОВ С ПРИМЕНЕНИЕМ МАШИННОГО ЗРЕНИЯ НА БАЗЕ УСТРОЙСТВ ИНТЕРНЕТА ВЕЩЕЙ

Создание проектов с применением нейронных сетей глубокого обучения часто связано с решением ряда технических задач: выбор сенсоров для получения качественных входных данных при умеренных затратах, оптимизация использования вычислительных ресурсов аппаратной платформы (особенно, если речь идёт об устройствах Интернета вещей – IoT), подготовка качественных обучающих датасетов и реализация их возможностей. **Цели исследования** – формулирование принципов решения перечисленных задач и разработка метода повышения качества распознавания для случаев, когда базовые решения достигли максимума своей эффективности. В данной статье это будет продемонстрировано на примере проекта Sortomat – автомата по приёму у населения перерабатываемой упаковки от различных товаров. **Методика исследования** базируется на применении воссоздаваемых результатов в ходе серии экспериментов и при математическом моделировании. **Результатом** является модификация автомата, которая с точки зрения **практической значимости** позволяет, сохраняя общую стоимость технического проекта, улучшать ряд характеристик Sortomat. Так, в качестве сенсоров в Sortomat выбраны камеры, позволяющие получать видеосигнал. Это значительно дешевле применения более сложных устройств анализа объекта, но позволяет добиться в большинстве случаев аналогичного качества распознавания. Для оптимизации использования вычислительных ресурсов применяются оптимизатор Adam и циклическое обучение. Это позволило уменьшить вес файла, описывающего веса, в два раза без потери качества распознавания (с сохранением уровня количества ошибок первого и второго порядка). При создании датасетов применяются алгоритм поиска области интереса на изображении и мажоритарная система распознавания объекта. Область интереса позволяет значительно снизить объём «бесполезной» информации на входах систем. Это повышает общее качество выборки (присутствуют только искомые объекты). Мажоритарная система позволяет принимать решения в ситуациях, когда объект оказывается похож на два и более объектов из разных классов. Она, методом весовых коэффициентов, определяет достоверность результатов по каждому из датасетов и выносит общий вердикт о принадлежности объекта к классу.

Ключевые слова: Интернет вещей, нейронная сеть глубокого обучения, машинное зрение, область интереса.

A.I. Tur, A.N. Kokoulin

Perm National Research Polytechnic University, Perm, Russian Federation

ISSUES OF OPTIMIZING OF PROJECTS COMPUTER VISION ON THE BASIS ON DEVICES INTERNET OF THINGS

Creating projects using deep learning neural networks is often associated with solving a number of technical problems: choosing sensors to obtain high-quality input data at a moderate cost, optimizing the use of hardware platform computing resources (especially when it comes to Internet of Things devices - IoT), preparing high-quality training datasets and realizing their capabilities. The **purpose** of the study is to formulate principles for solving these problems and to develop a method for improving the quality of recognition for cases when basic solutions have reached the maximum of their effectiveness. In this article, this will be demonstrated on the example of the Sortomat project - a machine for receiving processed packaging from the population from various goods. The **research methodology** is based on the use of recreated results during a series of experiments and mathematical modeling. The **results** are a modification of the machine, which, in terms of **practical significance**, allows, while maintaining the total cost of the technical project, to improve a number of Sortomat characteristics. Cameras are selected as sensors in it, allowing to receive a video signal. This is much cheaper than using more complex devices for analyzing an object, but it allows achieving, in most cases, a similar quality of recognition. The Adam optimizer and loop learning are used to optimize the use of computational resources. This made it possible to reduce the weight of the file describing the weight by half without losing the quality of recognition (while maintaining the level of the number of errors of the first and second order). When creating datasets, an algorithm for searching for a region of interest in an image and a majority object recognition system are used. The region of interest can significantly reduce the amount of "useless" information at the input of systems. This improves the overall quality of the sample (only the desired objects are present). The majority system allows making decisions in situations when an object is similar to two or more different classes of objects. Using the method of weighting factors, it determines the reliability of the results for each of the datasets and makes a general verdict on the belonging of the object to the class.

Keywords: internet of things, deep learning neural network, machine vision, region of interest.

Введение

В последние годы данные часто называют нефтью и золотом XXI века. Поскольку потребители, предприятия и другие организации генерируют данные с беспрецедентной скоростью, предприятия ищут эффективные способы анализа данных и получения информации, которая могла бы помочь им улучшить свои бизнес-процессы и оптимизировать принятие решений. В этом контексте они все чаще разрабатывают и внедряют системы и технологии искусственного интеллекта (ИИ), которые помогают им автоматизировать бизнес-процессы и принимать более обоснованные бизнес-решения. В течение последнего десятилетия интерес к машинному обучению (ML) в основном вращался вокруг глубокого обучения (DL) [1–4, 8], особого сегмента ML, который фокусируется на глубоких нейронных сетях (DNN). DNN – это нейронные сети с несколькими слоями нейронов между входом и выходом.

Приложения машинного обучения работают на различных типах устройств благодаря инструментам и методам, которые позволяют разрабатывать и развертывать модели машинного обучения на узлах с ограниченными ресурсами – устройств Интернета вещей (IoT) [5, 8]. В последнем случае приложения машинного обучения удобно называть устройствами AIoT (искусственный интеллект в IoT). Однако в реализации подобных решений имеются проблемы [6–8].

Reverse vending machine (RVM) – это автомат, позволяющий людям возвращать использованную тару для напитков на переработку за вознаграждение. Обычно автомат устроен таким образом, чтобы возвращать конечному пользователю сумму, эквивалентную стоимости сданной тары. Машины распознают каждый материал собранной тары, сортируют и уплотняют их в «пакеты», готовые к переработке [9]. Благодаря компактности и чистоте фракции, получаемой в результате такого подхода, стоимость каждого пакета выше, чем неуплотнённого несортированного мусора такого же объёма. На примере данного проекта будут рассмотрены основные проблемы схожих решений.

Проблемы реализации систем машинного зрения

Самая сложная задача при создании такого автомата – распознавание материала. Для этого, как правило, используют следующие методы:

– контроль материала контейнера (например, с помощью ИК-спектрометра);

– контроль формы контейнера;

– контроль штрихкода.

Эти три основные процедуры контроля позволяют почти гарантированно распознавать материал, но являются достаточно дорогими в плане реализации и эксплуатации. Благодаря современным технологиям компьютерного зрения мы смогли разработать еще один вид эффективных и недорогостоящих RVM, имеющих схожую функциональность. Данный проект называется Sortomat. Особенностью нашего проекта является то, что он основан на распознавании объектов нейронными сетями, используя одноплатный компьютер IoT – Raspberry PI 4. Автомат оснащён экраном, камерой и некоторыми датчиками, позволяющими быстро и качественно взаимодействовать с пользователем. Преимущество такого подхода заключается в том, что он не зависит от обновления баз данных штрихкодов, ему неважно наличие неповреждённой упаковки

(автомат принимает упаковку без этикеток, имеющую загрязнение, с допустимой деформацией формы), и, главное, не требуется идеальных условий для полноценной работы распознающего устройства. Нейронная сеть обучается на фотографиях банок, PET-бутылок, HDPE-контейнеров и определяет основные функции, необходимые для классификации. Благодаря этой особенности любое новое изображение может быть идентифицировано как совокупный класс, если контейнер обладает соответствующими признаками, что допускает создание полностью автономного устройства при использовании данного подхода.

Следующей проблемой в реализации подобного автомата является оптимизация такого процесса распознавания. На одноплатном компьютере запускается нейронная сеть, которая на основе поступающей входной информации с видеокамер принимает решение. Однако такой компьютер обладает ограниченной вычислительной мощностью, что чаще всего вызывает сбои в работе программы или достаточно заметные замедления процесса распознавания объекта. По этой причине к нейронной сети, помимо высокой точности, необходимо предъявлять дополнительные требования – высокое быстродействие и малый объем занимаемой памяти. Проблема заключается в преобразовании обученной нейронной сети таким образом, чтобы ее быстродействие увеличилось, размер модели нейронной сети уменьшился, а точность при этом осталась прежней или уменьшилась незначительно. Основываясь на ранее проведённых исследованиях по выбору архитектуры нейронной сети, в автомате используются два типа сетей – MobileNet и LeNet, которые относятся к сетям глубокого обучения и дают наибольшую точность среди других рассмотренных архитектур.

Для обучения использует уникальный набор изображений, полученный непосредственно автоматом Sortomat. База изображений разделена по материалу предмета и категории (пищевой контейнер, химический контейнер и т.д.). Вся база фотографий поделена на обучающую и тестовую выборки в соотношении 80/20. Для обучения используются облачное вычисление Cloud TPU и фреймворк TensorFlow. Тензорный процессор (TPU) Google – интегральная схема специального назначения (ASIC), разработанная компанией Google для выполнения задач по машинному обучению. Он работает в основных продуктах Google, включая Translate, Photos, Search Assistant и Gmail. Преимущества облачного TPU связаны с масштабируемостью и лёгкостью использования для всех

разработчиков и специалистов по изучению данных, запускающих передовые модели машинного обучения в облаке Google. TensorFlow – открытая программная библиотека для машинного обучения, разработанная также компанией Google для решения задач построения и тренировки нейронной сети с целью автоматического нахождения и классификации образов с достижением качества человеческого восприятия. Основной язык для работы с библиотекой реализован для Python, а также существуют реализации для R, C Sharp, C++, Haskell, Java, Go и Swift. Для выполнения задач непосредственно в автомате Sortomat применяется TensorFlow Lite – это платформа машинного обучения с открытым исходным кодом, позволяющая использовать TensorFlow в IoT и мобильных устройствах за счёт оптимизированных и облегчённых алгоритмов и принципов работы с памятью.

Для оптимизации нейронной сети используются оптимизатор Adam и циклическое обучение. Нейронная сеть на базовой тестовой выборке достигла общей точности 97 %, а по отдельно взятым классам превышает 99 %. Были проведены эксперименты по выбору метода квантования нейронной сети для уменьшения занимаемой памяти с минимальными потерями точности и скорости. Среди нескольких типов квантования (квантование динамического диапазона, квантование с репрезентативным набором данных, целочисленное квантование с репрезентативным набором данных, квантование во float16) выбран тип квантования float16, поскольку с его помощью файл весовых коэффициентов уменьшается на 50 %, а точность нейронной сети и скорость остаются прежними.

Однако, насколько бы эффективно нейронная сеть не распознавала базовый набор изображений, всегда будут находиться такие примеры, которые будут определяться неправильно – примеры, находящиеся на границе классов (рис. 1). В представленном случае достаточно наглядно видно, что есть объекты, значительно отличающиеся друг от друга (в понимании нейронной сети) – они разнесены в разные углы диаграммы. Однако имеются и примеры, которые достаточно похожи друг на друга – HDPE-контейнеры с вытянутым горлышком и матовые белые PET-бутылки. Часть таких объектов оказывается при тестировании распознана неправильно – красные и оранжевые точки.

Это является основополагающей проблемой состязательного обучения (англ. Adversarial training). Для решения этой проблемы часто

предлагают расширять обучающую выборку за счёт именно таких пограничных примеров [10]. Недостаток данного метода достаточно очевиден: на каждой итерации обучения, на каждый пример, мы можем сгенерировать очень большое количество примеров, соответственно, и время на обучение модели возрастает многократно, что противоречит предшествующему пункту. Кроме того, при чрезмерном включении таких примеров в выборку значительно снижается чувствительность нейронной сети по признакам базовых примеров, размывая границы между классами. Для решения данной проблемы мы предлагаем использовать области интереса (англ. Region of Interest, ROI) и группы нейронных сетей, управляемых мажоритарным способом.

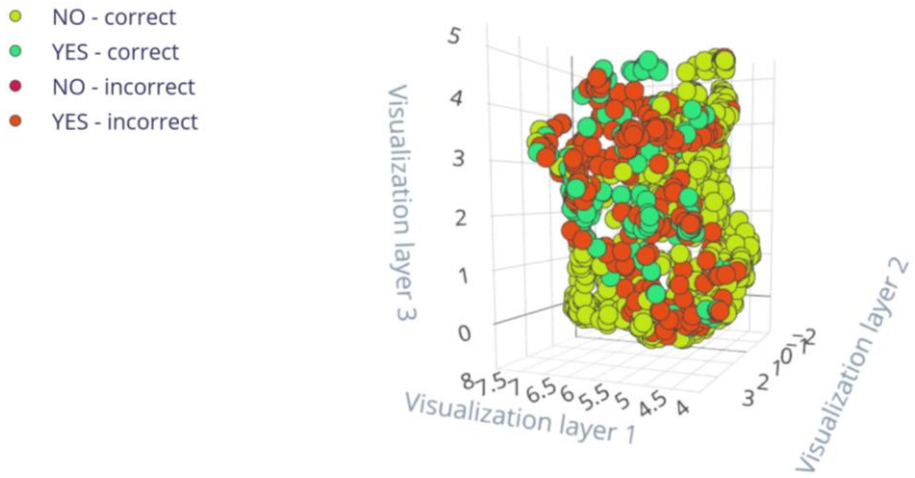

Рис. 1. Совокупность объектов классов PET и HDPE

Принцип работы ROI [11–13] обусловлен целью убрать все лишние данные, оставляя только необходимые для обработки. При распознавании изображения ставится цель обнаружить интересующий объект на изображении и вырезать его из остального фона. В нашем случае алгоритму необходимо будет ограничить область изображения, подаваемую на вход нейронной сети.

В электронике мажоритарным элементом (мажоритарным клапаном) называют логический элемент из класса пороговых, с чётным или нечётным числом входов и одним выходным сигналом, значение которого совпадает со значением на большинстве входов. При чётном числе входов большинством считается «n/2+1», соответственно, «n/2» к большинству не относится. Для автомата предложено использовать

аналогичный подход: если большинство сетей уверено, что на изображении бутылка из класса PET, то это решение и будет принято как результирующее.

Используя эти два подхода, процесс распознавания можно разбить на два этапа. Первый этап выполняется системой грубого поиска. Её основной задачей является обнаружение объекта и выделение области интереса с помощью быстрых алгоритмов машинного зрения. Применение сложных алгоритмов и нейронных сетей на данном этапе возможно, если процесс сортировки не имеет жёстких временных ограничений. На этом этапе должны выполняться все действия, направленные на повышение качества изображения перед конечным этапом сортировки: проверка наличия объекта в отсеке приема и определение его ROI, обрезка изображения по границам ROI (рис. 2), проверка расфокусировки и смазанности изображения [14–25].

Рис. 2. Пример нахождения ROI и обрезка изображения по заданным границам

В качестве датчиков можно использовать как обычные видеокамеры для получения изображения отсека, так и дополнительные – сенсоры ближнего или дальнего инфракрасного диапазона. Они позволяют игнорировать случайное загрязнение отсека, отсекать возможные блики и т.п. Этот этап позволяет получать более чистые входные данные нейронной сети, отсекая излишнюю информацию, способную дать ложные признаки для класса (рис. 3).

Второй этап – точный поиск. Блок нейронных сетей для точного поиска может состоять из одной или нескольких сетей. Для случая единственной нейронной сети обучение должно показывать максимально достижимый результат. Главным преимуществом такого варианта являются малая ресурсозатратность и относительно высокая скорость обработки. Это достигается за счёт предварительной подготовки изображения на первом этапе. Однако данный метод часто

будет страдать от случаев, когда признаки классов будут совпадать, – такие объекты могут быть идентифицированы неправильно. Параллельное использование нескольких нейронных сетей позволяет снизить влияние этого фактора на конечный результат, но при этом значительно повышает ресурсозатратность и может негативно влиять на общее время распознавания объекта.

Рис. 3. Пример изображений, поступающих на вход нейронной сети напрямую с камеры и с обрезкой по контурам ROI

Суть метода, как было отмечено выше, заключается в реализации согласованной работы некоторого количества нейронных сетей. Принципы и методы обучения при постановке этого эксперимента не имеют значения, так как не будут оказывать прямого влияния на предлагаемый метод. Для оценки возможностей сетей необходимо провести серию тестовых распознаваний идентичных объектов для получения начальной статистики их эффективности.

Допустим, что в нашем распоряжении имеются три варианта обученной нейронной сети (табл. 1, 2 и 3):

Таблица 1

Результаты обучения нейронной сети на первом датасете

Первый датасет	Ошибок	Точность,%	Ошибочно PET	Ошибочно HDPE	Ошибочно ALU
PET	143	52	–	94	49
HDPE	35	88	35	–	0
ALU	13	96	8	5	–
Всего	191	79	43	99	49

Таблица 2

Результаты обучения нейронной сети на втором датасете

Второй датасет	Ошибок	Точность, %	Ошибочно PET	Ошибочно HDPE	Ошибочно ALU
PET	83	72	–	71	12
HDPE	152	49	147	–	5
ALU	8	97	8	0	–
Всего	243	73	155	71	17

Таблица 3

Результаты обучения нейронной сети на третьем датасете

Третий датасет	Ошибок	Точность, %	Ошибочно PET	Ошибочно HDPE	Ошибочно ALU
PET	75	75	–	63	12
HDPE	51	83	36	–	15
ALU	36	88	25	11	–
Всего	162	82	61	74	27

Они имеют как свои плюсы, так и минусы, общая точность распознавания выше у третьего варианта (82 %), однако сеть чаще ошибается в определении классов HDPE и ALU по сравнению со вторым вариантом. Эти данные получены при анализе тестовой выборки, состоящей из 900 изображений (300 уникальных изображений каждого из классов). Анализ проводился для каждого из трёх вариантов нейронных сетей. Результаты проанализированы как по общим параметрам, так и по каждому классу в отдельности.

Как видно из результатов, ни одна обученная нейронная сеть не может показать достаточно качественного распознавания объекта и часто путает классы. Это происходит прежде всего из-за того, что тестовая выборка содержит фотографии реальных объектов, которые деформируются пользователями перед сдачей в автомат (рис. 4), а также объектов, изначально достаточно похожих друг на друга (рис. 5). Добиться идеального распознавания в подобных случаях сложно, так как часть признаков классов имеет много общего и может быть не воспринята даже человеком.

Рис. 4. Ошибочно распознанные деформированные
объекты класса PET и ALU

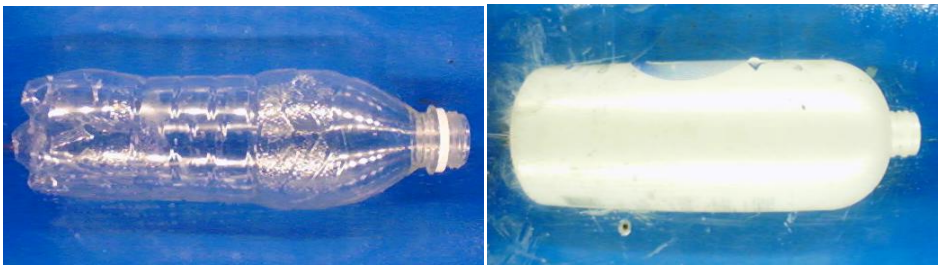

Рис. 5. Ошибочно распознанные похожие объекты класса PET и HDPE

Для правильной совместной работы нейронных сетей определяются веса, контролирующие степень доверия к результатам распознавания. Они могут быть выбраны равновесно (в зависимости от количества нейронных сетей) или асимметрично (в зависимости от точности распознавания каждой конкретной сети). В табл. 4 приведён пример асимметричных весов для рассмотренных выше нейронных сетей, рассчитанный, исходя из их точности определения каждого из классов:

$$f_{i,j} = \frac{N_i - R_{i,j}}{\sum_{j=1}^{3} N_i - R_{i,j}}, \quad (1)$$

где i – класс, j – номер нейронной сети, N – количество изображений класса в тестовой выборке, R – количество ошибок при распознавании класса нейронной сетью. Таким образом, в случае, если первая и третья нейронные сети сообщат, что объект принадлежит классу PET, а вторая – HDPE, то окончательное решение будет следующим: PET (0,26 + 0,38 > 0,22). Подобное решение значительно уменьшает чувствительность системы в целом к пересекающимся классам и позволяет обучать нейронные сети лишь с упором на один конкретный класс. Результаты эксперимента представлены в табл. 5 и на рис. 6.

Таблица 4

Вес результата распознавания

Класс	Первый датасет	Второй датасет	Третий датасет
PET	0,26	0,36	0,38
HDPE	0,40	0,22	0,38
ALU	0,34	0,35	0,31

Таблица 5

Результаты работы мажоритарной системы

Маж. система	Ошибок	Точность, %	Ошибочно PET	Ошибочно HDPE	Ошибочно ALU
PET	83	72	–	63	12
HDPE	46	85	36	–	15
ALU	9	97	25	11	–
Всего	138	85	61	74	27

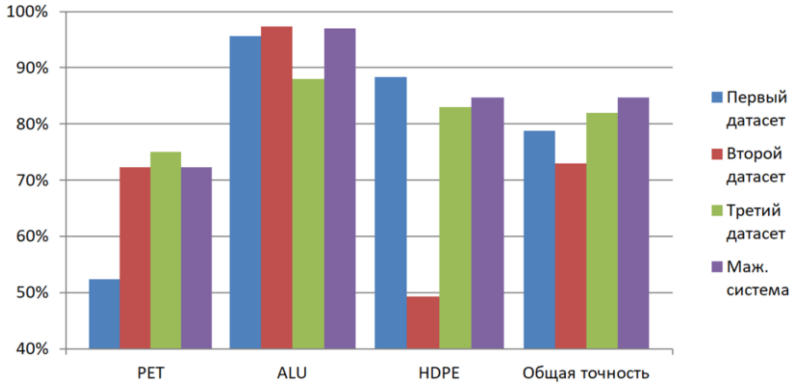

Рис. 6. Сопоставление результатов распознавания классов отдельными нейронными сетями и мажоритарной системой

В данном частном случае наблюдается компенсация в «провалах» эффективности распознавания каждой отдельной сети за счёт подключения сетей, более правильно распознающих конкретный класс. В соответствии с этим общая эффективность распознавания тоже увеличивается.

Применение данного метода при накоплении изображений сдаваемой тары позволяет создавать датасеты лучшего качества (по сравнению с накоплением обычных изображений, полученных от камеры).

Они позволяют избавиться от таких основных недостатков, как ложные срабатывания при изменении освещения в отсеке (чаще всего из-за бликов), ошибки при тренировке, связанные с «заученностью сцены», приводящие к переобучению.

Заключение

В ходе исследования был сформирован набор методов и подходов, позволяющих значительно повысить качество работы автомата Sortomat и снизить издержки, связанные с плохой оптимизацией системы принятия решения. Аналогичные шаги могут быть предприняты и к схожим проектам, опирающимся на теже принципы функционирования. Новизна полученных данных заключается в выдвижении набора современных методов, позволяющих при их совместном использовании улучшить технические характеристики систем распознавания, достигших своего аппаратного предела. Однако данные подходы могут значительно усложнять техническую реализацию проекта, так как повышают количество предварительной подготовки к обучению и сложность программной составляющей проекта.

В перспективе предполагаются следующие направления дальнейших исследований в рамках рассматриваемой тематики:

1) повышение количества одновременно используемых нейронных сетей в рамках ограниченных вычислительных мощностей;

2) исследование принципов искусственной генерации данных для обучения и их влияния на предложенные мажоритарные системы распознавания;

3) использование других методов оптимизации вычислений нейронной сети – кластеризация весовых коэффициентов и сокращение избыточных нейронов.

Исследование выполнено при финансовой поддержке правительства Пермского края в рамках научного проекта № С26/174.6.

Библиографический список

1. Shrestha A., Mahmood A. Review of Deep Learning Algorithms and Architectures // IEEE Access. – 2019. – Vol. 7. – P. 53040–53065. DOI: 10.1109/ACCESS.2019.2912200

2. Efficient Embedded Machine Learning applications using Echo State Networks / L Cerina, M.D. Santambrogio, G. Franco, C. Gallicchio, A. Micheli // Design, Automation & Test in Europe Conference & Exhibition (DATE). – 2020. – P. 1299–1302. DOI: 10.23919/DATE48585.2020.9116334

3. Roberto Saracco. TinyML: a glimpse into a future of Massive Distributed AI // IEEE Future Directions. – January 2021. – URL: https://cmte.ieee.org/futuredirections/2021/01/25/tinyml-a-glimpse-into-a-future-of-massive-distributed-ai/ (дата обращения: 04.04.2021).

4. Ibrahim A., Valle M. Real-Time Embedded Machine Learning for Tensorial Tactile Data Processing // IEEE Transactions on Circuits and Systems I: Regular Papers. – Nov. 2018. – Vol. 65, no. 11. – P. 3897–3906. DOI: 10.1109/TCSI.2018.2852260

5. Powering the IoT through embedded machine learning and LoRa / V.M. Suresh, R. Sidhu, P. Karkare, A. Patil, Z. Lei, A. Basu // 4th World Forum on Internet of Things (WF-IoT). – Singapore, 2018. – P. 349–354. DOI: 10.1109/WF-IoT.2018.8355177

6. Integrating machine learning in embedded sensor systems for Internet-of-Things applications / J. Lee, M. Stanley, A. Spanias, C. Tepedelenlioglu // International Symposium on Signal Processing and Information Technology (ISSPIT). – Limassol, 2016. – P. 290–294. DOI: 10.1109/ISSPIT.2016.7886051

7. Andrade L., Prost-Boucle A., Pétrot F. Overview of the state of the art in embedded machine learning // Design, Automation & Test in Europe Conference & Exhibition (DATE). – Dresden, 2018. – P. 1033–1038, DOI: 10.23919/DATE.2018.8342164

8. Soldatos John. The Embedded Machine Learning Revolution: The Basics You Need to Know // Информационный портал Wevolver. – URL: https://www.wevolver.com/article/the-embedded-machine-learning-revolution-the-basics-you-need-to-know (дата обращения: 04.04.2021).

9. Beverage Container Collecting Machine Project / A.I. Tur, A.N. Kokoulin, A.A. Yuzhakov, S.V. Polygalov, A.S. Troegubov, V.N. Korotaev // IOP Conf. Series: Earth and Environmental Science. – 2019. – Vol. 317. – Art. 012006. – 9 p.

10. Szegedy Christian. Intriguing properties of neural networks. – URL: https://arxiv.org/pdf/1312.6199.pdf

11. Hierarchical Convolutional Neural Network Architecture in Distributed Facial Recognition System / A.I. Tur, A.N. Kokoulin,

A.A. Yuzhakov, A.I. Knyazev // Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). – 2019. – P. 1–5.

12. Akhmetzyanov K.R., Yuzhakov A.A. Waste Sorting Neural Network Architecture Optimization // Proceedings 2019 International Russian Automation Conference (RusAutoCon). – 2019. – P. 1–5.

13. Kokoulin A.N., Kiryanov D.A. The Optical Subsystem for the Empty Containers Recognition and Sorting in a Reverse VendingMachine // 4th International Conference on Smart and Sustainable Technologies, SpliTech. – 2019 (статья No 8782990).

14. Kokoulin, A.N.,Yuzhakov, A.A.,Tur, A.I.,Knyazev, A.I. The optical method for the plastic waste recognition and sorting in a reverse vending machine // International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM. – 2019. – 19(4.1). – P. 793–800.

15. Kokoulin A., May I., Kokoulina A. Image Processing Methods in Analysis of Component Composition and Distribution of Dust Emissions for Environmental Quality Management // Proceedings of 10th International Conference on Large-Scale Scientific Computations (LSSC); Bulgarian Acad Sci, Sozopol, BULGARIA. – 2015. – Jun. 08–12. – Vol. 9374. – P. 352–359.

16. Supriya Suresh & Subaji Mohan. ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis // Neural Computing and Applications. – 2020.

17. Тур А.И., Кокоулин А.Н., Дзыгарь А.В. Иерархическая система поиска и распознавания штрихкода на повреждённой таре в автомате раздельного сбора отходов // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2019. – № 29. – С. 44–57.

18. Clay D. Spence, John C. Pearson, Jim Bergen. Coarse-to-Fine Image Search Using Neural Networks. – URL: https://papers.nips.cc/paper/982-coarse-to-fine-image-search-using-neural-networks.pdf

19. Cheng Lei, Yee-Hong Yang. Optical Flow Estimation on Coarse-to-Fine Region-Trees using Discrete Optimization. – URL: https://cs.brown.edu/courses/cs296-4/Papers/2010/iccv2009_201.pdf

20. Кулаков И.Ю., Вологин Д.А., Пикалов В.В. Многосеточный алгоритм в задаче веерной ROI-томографии // Теория и численные методы решения обратных и некорректных задач: материалы V Междунар. молодеж. науч. шк.-конф. (Новосибирск, 8–13 октября 2013 г.). – Новосибирск, 2013.

21. End to end learning for self-driving cars / M. Bojarski [et al.] // arXiv preprint arXiv:1604.07316. – 2016.

22. Bekey G.A., Goldberg K.Y. (eds.). Neural networks in robotics // Springer Science & Business Media. – 2012. – Vol. 202.

23. Quantization and training of neural networks for efficient integer-arithmetic-only inference / B. Jacob [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – P. 2704–2713.

24. Han S., Mao H., Dally W.J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding // arXiv preprint arXiv:1510.00149. – 2015.

25. Low-bit quantization of neural networks for efficient inference / Y. Choukroun [et al.] // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE. – 2019. – P. 3009–3018.

References

1. Shrestha A., Mahmood A. Review of Deep Learning Algorithms and Architectures. *IEEE Access*, 2019, vol. 7, pp. 53040-53065. DOI: 10.1109/ACCESS.2019.2912200

2. Cerina L. Santambrogio M.D., Franco G., Gallicchio C., Micheli A. Efficient Embedded Machine Learning applications using Echo State Networks. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2020, pp. 1299-1302. DOI: 10.23919/DATE48585.2020.9116334

3. Roberto Saracco. TinyML: a glimpse into a future of Massive Distributed AI. *IEEE Future Directions.* January 2021, available at: https://cmte.ieee.org/futuredirections/2021/01/25/tinyml-a-glimpse-into-a-future-of-massive-distributed-ai/ (accessed 04 April 2021).

4. Ibrahim A., Valle M. Real-Time Embedded Machine Learning for Tensorial Tactile Data Processing. *IEEE Transactions on Circuits and Systems I: Regular Papers*, November 2018, vol. 65, no. 11, pp. 3897-3906. DOI: 10.1109/TCSI.2018.2852260

5. Suresh V.M., Sidhu R., Karkare P., Patil A., Lei Z., Basu A. Powering the IoT through embedded machine learning and LoRa. *4th World Forum on Internet of Things (WF-IoT).* Singapore, 2018, pp. 349-354. DOI: 10.1109/WF-IoT.2018.8355177

6. Lee J., Stanley M., Spanias A., Tepedelenlioglu C. Integrating machine learning in embedded sensor systems for Internet-of-Things applications. *International Symposium on Signal Processing and Information Technology (ISSPIT)*. Limassol, 2016, pp. 290-294. DOI: 10.1109/ISSPIT.2016.7886051.

7. Andrade L., Prost-Boucle A., Pétrot F. Overview of the state of the art in embedded machine learning. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*. Dresden, 2018, pp. 1033-1038, DOI: 10.23919/DATE.2018.8342164

8. Soldatos John. The Embedded Machine Learning Revolution: The Basics You Need to Know. *Informatsionnyi portal Wevolver*, available at: https://www.wevolver.com/article/the-embedded-machine-learning-revolution-the-basics-you-need-to-know (accessed 04 April 2021).

9. Tur A.I., Kokoulin A.N., Yuzhakov A.A., Polygalov S.V., Troegubov A.S. Korotaev V.N. Beverage Container Collecting Machine Project. *IOP Conf. Series: Earth and Environmental Science*, 2019, vol. 317, Art. 012006, 9 p.

10. Christian Szegedy. Intriguing properties of neural networks, available at: https://arxiv.org/pdf/1312.6199.pdf

11. Tur A.I., Kokoulin A.N., Yuzhakov A.A., Knyazev A.I. Hierarchical Convolutional Neural Network Architecture in Distributed Facial Recognition System. *Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 2019, pp. 1-5.

12. Akhmetzyanov K.R., Yuzhakov A.A. Waste Sorting Neural Network Architecture Optimization. *Proceedings 2019 International Russian Automation Conference (RusAutoCon)*, 2019, pp. 1-5.

13. Kokoulin A.N., Kiryanov D.A. The Optical Subsystem for the Empty Containers Recognition and Sorting in a Reverse Vending Machine. *4th International Conference on Smart and Sustainable Technologies*, SpliTech, 2019.

14. Kokoulin A.N., Yuzhakov A.A., Tur A.I., Knyazev A.I. The optical method for the plastic waste recognition and sorting in a reverse vending

machine. *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, 2019, 19 (4.1), pp. 793-800.

15. Kokoulin A., May I., Kokoulina A. Image Processing Methods in Analysis of Component Composition and Distribution of Dust Emissions for Environmental Quality Management. *Proceedings of 10th International Conference on Large-Scale Scientific Computations (LSSC)*; Bulgarian Acad Sci, Sozopol, BULGARIA, June 08-12 2015, vol. 9374, pp. 352-359.

16. Supriya Suresh & Subaji Mohan. ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis. *Neural Computing and Applications*, 2020.

17. Tur A.I., Kokoulin A.N., Dzygar' A.V. Ierarkhicheskaia sistema poiska i raspoznavaniia shtrikhkoda na povrezhdennoi tare v avtomate razdel'nogo sbora otkhodov [Hierarchical system for searching and recogniz-ing a barcode on damaged containers in a separate waste collection machine]. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniia*, 2019, no. 29, pp. 44-57.

18. Clay D. Spence, John C. Pearson, Jim Bergen. Coarse-to-Fine Image Search Using Neural Networks, available at: https://papers.nips.cc/paper/982-coarse-to-fine-image-search-using-neural-networks.pdf

19. Cheng Lei, Yee-Hong Yang. Optical Flow Estimation on Coarse-to-Fine Region-Trees using Discrete Optimization, available at: https://cs.brown.edu/courses/cs296-4/Papers/2010/iccv2009_201.pdf

20. Kulakov I.Iu., Vologin D.A., Pikalov V.V. Mnogosetochnyi algoritm v zadache veernoi ROI-tomografii [Multigrid algorithm in the problem of fan ROI tomography]. *Teoriia i chislennye metody resheniia obratnykh i nekorrektnykh zadach: V Mezhdunarodnaia molodezhnaia nauchnaia shkola-konferentsiia* (Novosibirsk, 8-13 October 2013). Novosibirsk, 2013.

21. Bojarski M. et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

22. Bekey G.A., Goldberg K.Y. (eds.). Neural networks in robotics. *Springer Science & Business Media*, 2012, vol. 202.

23. Jacob B. et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2704-2713.

24. Han S., Mao H., Dally W.J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149*, 2015.

25. Choukroun Y. et al. Low-bit quantization of neural networks for efficient inference. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW).* IEEE, 2019, pp. 3009-3018.

Сведения об авторах

Тур Александр Игоревич (Пермь, Россия) – кандидат технических наук, ассистент кафедры «Автоматика и телемеханика» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: tur.aleksandr93@mail.ru).

Кокоулин Андрей Николаевич (Пермь, Россия) – кандидат технических наук, доцент кафедры «Автоматика и телемеханика» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: a.n.kokoulin@at.pstu.ru).

About the authors

Alexander I. Tur (Perm, Russian Federation) – Ph. D. in Technical Sciences, Assistant of the Department Automatic and Telemechanic Perm National Research Polytechnic University (614990, Perm, 29, Komsomolsky pr., e-mail: tur.aleksandr93@mail.ru).

Andrey N. Kokoulin (Perm, Russian Federation) – Ph. D. in Technical Sciences, Associate Professor of the Department Automatic and Telemechanic Perm National Research Polytechnic University (614990, Perm, 29, Komsomolsky pr., e-mail: a.n.kokoulin@at.pstu.ru).

Получено 08.04.2021