

На правах рукописи

АХМЕТЗЯНОВ КИРИЛЛ РАИСОВИЧ

**НЕЙРО-СЕТЕВЫЕ МЕТОДЫ И АЛГОРИТМЫ САМООБУЧЕНИЯ ПРИ
ОБРАБОТКЕ ДАННЫХ В СИСТЕМЕ АВТОМАТИЗАЦИИ ПРОЦЕССА
СОРТИРОВКИ БЫТОВЫХ ОТХОДОВ**

05.13.06 – Автоматизация и управление
технологическими процессами и производствами (в промышленности)

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Пермь 2021

Работа выполнена на кафедре «Автоматика и телемеханика» в ФГАОУ ВО «Пермский национальный исследовательский политехнический университет».

Научный руководитель:

Южаков Александр Анатольевич
доктор технических наук, профессор,
заведующий кафедрой автоматике и
телемеханики ФГАОУ ВО «Пермский
национальный исследовательский
политехнический университет»

Официальные оппоненты:

Катасёв Алексей Сергеевич
доктор технических наук, доцент,
профессор кафедры систем
информационной безопасности
ФГБОУ ВО «Казанский национальный
исследовательский технический
университет им. А.Н. Туполева-КАИ»

Скакун Анастасия Дмитриевна
кандидат технических наук, доцент,
доцент кафедры систем автоматического
управления ФГАОУ ВО «Санкт-
Петербургский государственный
электротехнический университет «ЛЭТИ»
им. В.И. Ульянова (Ленина)»

Ведущая организация:

ФГБОУ ВО «Томский государственный
университет систем управления и
радиоэлектроники», г. Томск

Защита состоится «10» декабря 2021 года в 16.00 часов на заседании диссертационного совета Пермского национального исследовательского политехнического университета Д ПНИПУ.05.04 по адресу: 614990, г. Пермь, Комсомольский проспект, 29, ауд.345.

С диссертацией можно ознакомиться в библиотеке и на официальном сайте Пермского национального исследовательского политехнического университета (<http://pstu.ru/>).

Автореферат разослан «8» октября 2021 года

Учёный секретарь диссертационного
совета Д ПНИПУ.05.04
доктор технических наук, доцент

/ В.И. Фрейман /

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования и степень ее разработанности.

В России ежегодно в среднем накапливается 3 млн.т пластиковых отходов, и это количество с годами увеличивается, а время разложения пластика может достигать 1000 лет. В то же время количество перерабатываемого пластика в стране составляет менее 7%. Существующие заводы по переработке пластика не справляются с огромным потоком мусора. Для переработки пластика необходимо отсортировать его по видам пластиковых отходов, например, пластиковые бутылки, так как у каждого пластика своя температура плавления. Поэтому важно в режиме «реального времени» как можно точнее определять эти предметы среди прочего мусора, которые состоят из пластика.

В настоящее время этап сортировки пластиковых материалов на предприятиях по переработке отходов является одним из наименее автоматизированных этапов технологической цепочки. Для увеличения пропускной способности подобных предприятий необходимо разработать аппаратурно-программные средства по сортировке пластиковых материалов.

В информационных технологиях определение предметов называется классификацией объектов. Классификация объектов — это алгоритмы и набор математических преобразований, которые позволяют определить (идентифицировать) предмет.

Значительный вклад в развитие «умных» городов и автоматизации процесса сортировки бытовых отходов внесли такие зарубежные ученые, как A.R. Al-Ali, H. Basri, D.A. Wahab, H. Fu и W. Yousef, так и отечественные ученые – М.В. Соколов, Б.Б. Бобович, О.Б. Ганин, С.В. Абламейко, А.П. Добрынин, С.Н. Максимов и др.

Ученые T. Raiko и N. Krishnan применили и адаптировали технологии машинного и глубокого обучения, а также компьютерного зрения в области сортировки отходов. T. Raiko является разработчиком автоматизированной системы сортировки отходов. Его технология заключается в фотографировании мусора с помощью видеокамеры, применении нейронных сетей и компьютерного зрения для сортировки отходов по категориям и материалу.

Существует множество методов компьютерного зрения для распознавания объектов. К классическим методам относятся SIFT, SURF, LBP и HOG. Эти методы обладают следующими преимуществами: достаточно высокое быстродействие, устойчивость к контрастности изображения, поворотам, масштабам и частичному закрытию объектов распознавания. К недостаткам относится то, что объекты без ярко выраженной текстуры и с фрактальной структурой будут неверно распознаваться.

Другим классом методов распознавания являются методы на основе сверточных нейронных сетей (наиболее известные LeNet, AlexNet, VGG, ResNet, MobileNet). Преимуществом этих методов является возможность аппроксимировать любую функцию и обучить модель классифицировать любые объекты. К недостаткам относится то, что необходимо проводить вычислительно затратную

процедуру поиска оптимальных гиперпараметров, которые для каждого набора обучающей выборки могут различаться.

В диссертационной работе исследуется задача распознавания объектов на основе изображений в условиях ограниченных вычислительных ресурсов при выборе оптимальных гиперпараметров. Для ее реализации предлагается использовать многозадачный многокритериальный метод гиперпараметрической оптимизации. Такой метод позволит выбрать гиперпараметры обучения математических моделей распознавания на основе нескольких заданных критериев (достигаемая максимальная точность распознавания и время обучения модели) и нескольких задач распознавания. Это дает возможность применения сверточных нейронных сетей при меньших требованиях к вычислительной платформе без снижения качества распознавания.

Объект исследования: автоматизированный процесс предварительной сортировки бытовых отходов для переработки на предприятиях отраслей промышленности.

Предмет исследования: научно-методический аппарат автоматической классификации изображений с помощью сверточных нейронных сетей.

Цель работы: совершенствование автоматизированной сортировки бытовых отходов по заданным критериям качества на основе разработки и внедрения методов и алгоритмов самообучения систем автоматизации процессов сортировки с применением сверточных нейронных сетей.

Задачи работы. Для достижения поставленной цели необходимо решить следующие задачи:

1) провести анализ принципов функционирования существующих систем автоматизированной сортировки бытовых отходов;

2) построить критерии оптимизации гиперпараметров для автоматизированной сортировки бытовых отходов;

3) разработать метод оптимизации сверточной нейронной сети с комплексом заданных гиперпараметров для повышения качества процесса обучения;

4) разработать метод оптимизации вычислительных затрат при построении модели автоматической классификации изображений бытовых отходов;

5) разработать метод автоматического обучения распознавания при сортировке бытовых отходов;

б) внедрить предложенные методы оптимизации гиперпараметров и вычислительных затрат в устройство по автоматизированной сортировке бытовых отходов.

Методы исследования основаны на математических и методических принципах построения нейронных сетей, теории машинного обучения, обработки изображений, распознавания образов, теории оптимизации, теории планирования и обработки экспериментальных данных.

Основные положения, выносимые на защиту:

1. Предложенные критерии для оптимизации гиперпараметров моделей классификации изображений (*п. 15. Теоретические основы, методы и алгоритмы*

интеллектуализации решения прикладных задач при построении АСУ широкого назначения (АСУТП, АСУП, АСТПП и др.).

2. Метод многозадачной многокритериальной гиперпараметрической оптимизации вычислений математической модели (*п. 15. Теоретические основы, методы и алгоритмы интеллектуализации решения прикладных задач при построении АСУ широкого назначения (АСУТП, АСУП, АСТПП и др.).*)

3. Метод оптимизации вычислительных затрат при построении модели автоматической классификации изображений бытовых отходов (*п. 15. Теоретические основы, методы и алгоритмы интеллектуализации решения прикладных задач при построении АСУ широкого назначения (АСУТП, АСУП, АСТПП и др.).*)

4. Метод автоматического обучения специализированной нейронной сети для сортировки мусора (*п. 8. Формализованные методы анализа, синтеза, исследования и оптимизация модульных структур систем сбора и обработки данных в АСУТП, АСУП, АСТПП и др.)*

5. Программный комплекс с учетом разработанных методов в составе устройства по автоматизированной сортировке бытовых отходов (*п. 15. Теоретические основы, методы и алгоритмы интеллектуализации решения прикладных задач при построении АСУ широкого назначения (АСУТП, АСУП, АСТПП и др.).*)

Научная новизна:

1. Состав критериев для многокритериальной оптимизации при обучении моделей классификации изображений, отличающийся тем, что он сформирован с учетом гиперпараметров обучения и позволяет повысить точность модели классификации и снизить затраты вычислительных ресурсов автоматизированной системы.

2. Новый метод многокритериальной оптимизации гиперпараметров нейронной сети, отличающийся многозадачностью, что позволяет получить оптимальные гиперпараметры для заданных с их учетом критериев.

3. Новый метод оптимизации вычислений на основе квантования, отличающийся заданием нескольких критериев оптимизации, что позволяет уменьшить размер модели классификации без потери ее точности.

4. Новый метод сортировки бытовых отходов, отличающийся автоматическим самообучением предложенной нейронной сети, что позволяет снизить вычислительные затраты при оперативном обучении сети на примерах (данных), представляемых организациями, производящие сортировку отходов.

5. Программный комплекс в составе устройства по автоматизированной сортировке бытовых отходов, включающий нейро-сетевые методы и алгоритмы самообучения, который не требует дополнительных капитальных вложений в оборудование системы и обеспечивает повышение качества сортировки бытовых отходов.

Теоретическая значимость:

Создание научно-методического аппарата, заключающегося в разработке критериев оптимизации гиперпараметров моделей классификации изображений,

построении новых методов многозадачной многокритериальной гиперпараметрической оптимизации, оптимизации вычислений на основе квантования и автоматического обучения специализированной нейронной сети для сортировки мусора. Разработанные методы и алгоритмы применимы в других областях, требующих автоматизации с применением методов компьютерного зрения.

Практическая значимость заключается:

1) в разработке и программной реализации автоматизированной системы по сортировке бытовых отходов на основе предложенных методов, что позволило уменьшить общее время обучения нейронной сети в условиях ограниченных вычислительных ресурсов на 10% при испытаниях (в «Сортомате»), при сохранении точности полученной нейронной сети; повысить эффективность используемых вычислительных ресурсов АСУТП для распознавания объекта на изображении на 15% при испытаниях (в «Сортомате»);

2) в возможности использования затратных эффективных алгоритмов распознавания объектов в промышленных АСУТП с ресурсными ограничениями на предприятиях по переработке бытовых отходов без потери точности распознавания;

3) в применимости разработанных методов и алгоритмов в подсистемах компьютерного зрения автоматизации производственных процессов без изменений разработанных методов и алгоритмов.

Достоверность и обоснованность результатов. Общие тенденции, полученные в результате исследования, не противоречат результатам, представленных в литературе другими исследователями, а также подтверждаются сопоставлением теоретических выводов с результатами имитационных экспериментов и результатами внедрения устройства по автоматизированной сортировке бытовых отходов «Сортомат».

Апробация работы. Основные результаты диссертационной работы представлялись и обсуждались на XXI международной конференции по мягким вычислениям и измерениям (Россия, г. Санкт-Петербург, 2018 г.), IV и V всероссийской научно-практической конференции «Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века» (Россия, г. Пермь, 2019 и 2020 г.), 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (Россия, г. Санкт-Петербург, 2020 г.), 2019 и 2020 International Russian Automation Conference (Россия, г. Сочи, 2019 и 2020 г.).

Работы по теме диссертационного исследования выполнялись в рамках научного проекта № С-26/174.6 международной исследовательской группы учёных (МИГ-30).

Публикации. Основные результаты диссертационной работы опубликованы в 14 научных работах, из них: 7 статей в журналах, входящих в перечень ведущих журналов и изданий, рекомендуемых ВАК; 3 в изданиях, индексируемых в базах SCOPUS; 1 патент на полезную модель и 1 свидетельство о регистрации программы для ЭВМ, остальные – в тезисах докладов, материалах конференций и прочих источниках.

Объем и структура работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы из 127 наименований и 8 приложений. Полный объем диссертации составляет 121 страница, из которых 96 страниц занимает основной текст диссертации, включающий 21 рисунок и 15 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** приведено описание методов и алгоритмов распознавания объектов по изображению. Описаны основные проблемы производственных комплексов автоматизации технологических процессов, возникающие при реализации системы распознавания объектов. Представлено обоснование актуальности данной работы, поставлена цель и сформулированы задачи. Определены объект и предмет исследования, выделены научная новизна и основные положения, выносимые на защиту.

В **первой главе** анализируются существующие датчики и алгоритмы для обработки информации для сортировки бытовых отходов. Среди рассмотренных датчиков (NIR-датчик, электростатическая сортировка, рентгеновский датчик, датчик в видимом диапазоне) особый интерес представляет датчик в видимом диапазоне из-за универсальности и низкой стоимости, что позволяет со временем пополнять базу данных и добавлять новые категории мусора.

Выдвинута задача самообучающейся подсистемы распознавания АСУТП (1):

$$\begin{cases} \mathbb{E}_{a_1 \in A_1} \text{cost}(a_1) < \mathbb{E}_{a_2 \in A_2} \text{cost}(a_2), \\ \mathbb{E}_{a_1 \in A_1} \text{acc}(a_1) > \mathbb{E}_{a_2 \in A_2} \text{acc}(a_2), \\ A_1 \supset A_2 \end{cases} \quad (1)$$

где A_1 и A_2 – множества подсистем распознавания АСУТП, cost – вычислительные затраты на создания подсистем распознавания АСУТП, acc – точность подсистемы распознавания АСУТП.

Анализ существующих алгоритмов для обработки информации в видимом диапазоне показал, что традиционными являются SIFT, SURF и HOG. Наибольшую популярность в последнее время набирают сверточные нейронные сети, способные извлекать сложную скрытую структурированную информацию из изображений, благодаря чему достигается высокая точность классификации изображений.

Показано, что предпочтительно использование датчиков в видимом диапазоне с применением нейро-сетевых технологий для распознавания отходов, а эффективное решение задачи распознавания в условиях ограниченного вычислительного ресурса требует построения специализированной архитектуры нейронной сети и подбора алгоритма обучения.

Во **второй главе** представлены результаты экспериментов по выбору сверточной нейронной сети как математического аппарата обработки изображения и по выбору способа преобразования исходных изображений с помощью аугментации для увеличения точности выбранной нейронной сети на заданном типе оборудования (в условиях ограничения требований к вычислительным ресурсам).

Выбор сверточных нейронных сетей производился среди AlexNet, SqueezeNet и MobileNet, так как у этих сетей скорость работы на RaspberryPi (представитель вычислительных ресурсов с ограниченными возможностями) удовлетворяет

поставленной задаче (время обработки меньше 1 секунды). Результаты эксперимента представлены в табл.1 и 2.

Таблица 1. Точность распознавания на проверочной выборке

	AlexNet		SqueezeNet		MobileNet	
	Один слой	Вся сеть	Один слой	Вся сеть	Один слой	Вся сеть
Процент правильного распознавания	97,84	98,44	98,11	98,64	97,85	98,39

Таблица 2. Время распознавания одного изображения

	AlexNet		SqueezeNet		MobileNet	
	Один слой	Вся сеть	Один слой	Вся сеть	Один слой	Вся сеть
Время/изображение на CPU (мс)	174	181	373	400	118	117
Время/изображение на RaspberryPi (мс)	887	900	323	317	748	725

Самый высокий процент правильного распознавания среди указанных в таблице 1, имеет сеть MobileNet при настройке всех слоев. Минимальное время, необходимое для обработки одного изображения на RaspberryPi среди времен, указанных в таблице 2, имеет сеть SqueezeNet.

Другой эксперимент позволил осуществить выбор способа аугментации для MobileNet среди следующих преобразований исходных изображений: поворот на случайный угол, добавление гауссовского шума, отступ, удаление части изображения, вырезание части изображения и масштабирование (см. табл. 3-8).

Таблица 3. Результаты проверки поворота

Угол поворота	15 градусов	30 градусов	45 градусов	60 градусов	75 градусов
% правильной классификации	68	66	75	78	65

Таблица 4. Результаты проверки гауссовского шума

Стандартное отклонение	0,05	0,10	0,15	0,20
% правильной классификации	62	64	66	61

Таблица 5. Результаты проверки отступа

Доля высоты и ширины	0,1	0,2	0,3	0,4
% правильной классификации	72	63	73	66

Таблица 6. Результаты проверки удаления части изображения

Доля площади	0,2	0,3	0,4
% правильной классификации	56	66	65

Таблица 7. Результаты проверки вырезания части изображения

Диапазон доли площади	От 0,9 до 1,0	От 0,7 до 0,9	От 0,5 до 0,7	От 0,3 до 0,5
% правильной классификации	72	68	62	60

Таблица 8. Результаты проверки масштабирования

Диапазон доли площади	От 0,7 до 0,9	От 0,5 до 0,7	От 0,3 до 0,5	От 0,1 до 0,3
% правильной классификации	63	68	71	65

Наибольшая точность классификации достигнута при применении поворота на случайный угол от 0° до 60° и равен 78%. То есть точность классификации пластиковых бутылок и алюминиевых банок увеличилась на 20%.

Экспериментальные исследования позволили выбрать в качестве основы нейронную сеть MobileNet и для нее увеличить точность с помощью метода аугментации.

В третьей главе проведено сравнение существующих методов оптимизации гиперпараметров и предложен метод многозадачной многокритериальной гиперпараметрической оптимизации (Multi-Task MultiCriteria – МТМС) и метод оптимизации вычислений нейронной сети.

Существующие методы оптимизации гиперпараметров различаются по следующим признакам:

- 1) количеству оптимальных решений;
- 2) количеству решаемых задач;
- 3) количеству критериев выбора оптимального решения.

Для заданного набора изображений решается задача оптимизации с количеством классов $N_{classes}$ и количеством изображений N_{images} . Между задачами есть примеры одних и тех же классов, отличие заключается в том, как создаются изображения (разное освещение, фон и используемые камеры). Критерий — это количественная характеристика обучения / оценки нейронной сети по задаче (например, точность, время обработки одного изображения или эпоха, на которой процесс обучения нейронной сети достигает сходимости).

Рассмотренные методы оптимизации гиперпараметров не позволяют производить оптимизации одновременно по разным критериям и по различным задачам. Поэтому разработан новый метод оптимизации гиперпараметров, который устраняет указанный недостаток.

Выбор оптимальных гиперпараметров формализуется следующим образом:

$$\theta = \underset{\theta \in \Theta}{\operatorname{argmin}} \mathbb{E}_\tau [L(\theta, \phi)] \quad (2)$$

где Θ – множество всех гиперпараметров, θ – выбранные оптимальные гиперпараметры, ϕ – вектор коэффициентов значимости критериев, $L(\cdot)$ – функция

оценки модели с заданными гиперпараметрами θ и коэффициентами ϕ , τ – задача, для которой выполняется оптимизация.

Предлагаемый метод дает решение задачи (2), при этом МТМС удовлетворяет следующим требованиям:

- решает задачу минимизации;
- значимость каждого критерия определяется вектором коэффициентов ϕ (чем выше коэффициент, тем важнее соответствующий критерий).

Рассмотрим реализацию предлагаемого МТМС.

Введем обозначение тестовой выборки задачи τ :

$$x^i \sim D, i = 1..N_{task} \quad (3)$$

где x^i – это i -я тестовая выборка, имеющая распределение D , N_{task} – количество задач.

Получаем оценочную матрицу для x^i и θ :

$$V = M(x^i; \theta) \quad (4)$$

$$M(x^i; \theta) :: (\mathbb{R}^{x_{size}}, \mathbb{R}^{N_{parameter} \times N_{combination}}) \rightarrow \mathbb{R}^{N_{combination} \times N_{criteria}} \quad (5)$$

где $M(\cdot)$ – функция модели, которая преобразует заданную выборку x^i и с заданными гиперпараметрами θ в матрицу оценки V , $N_{criteria}$ – количество критериев, x_{size} – размер тестовой выборки, $N_{parameter}$ – количество гиперпараметров, $N_{combination}$ – количество комбинаций гиперпараметров.

Определим L (данная функция вычисляется для каждой выборки x^i):

$$L(\cdot; \theta, \phi) = E(V; \phi) \quad (6)$$

$$E(V; \phi) :: (\mathbb{R}^{N_{criteria}}, \mathbb{R}^{N_{criteria}}) \rightarrow \mathbb{R}^1 \quad (7)$$

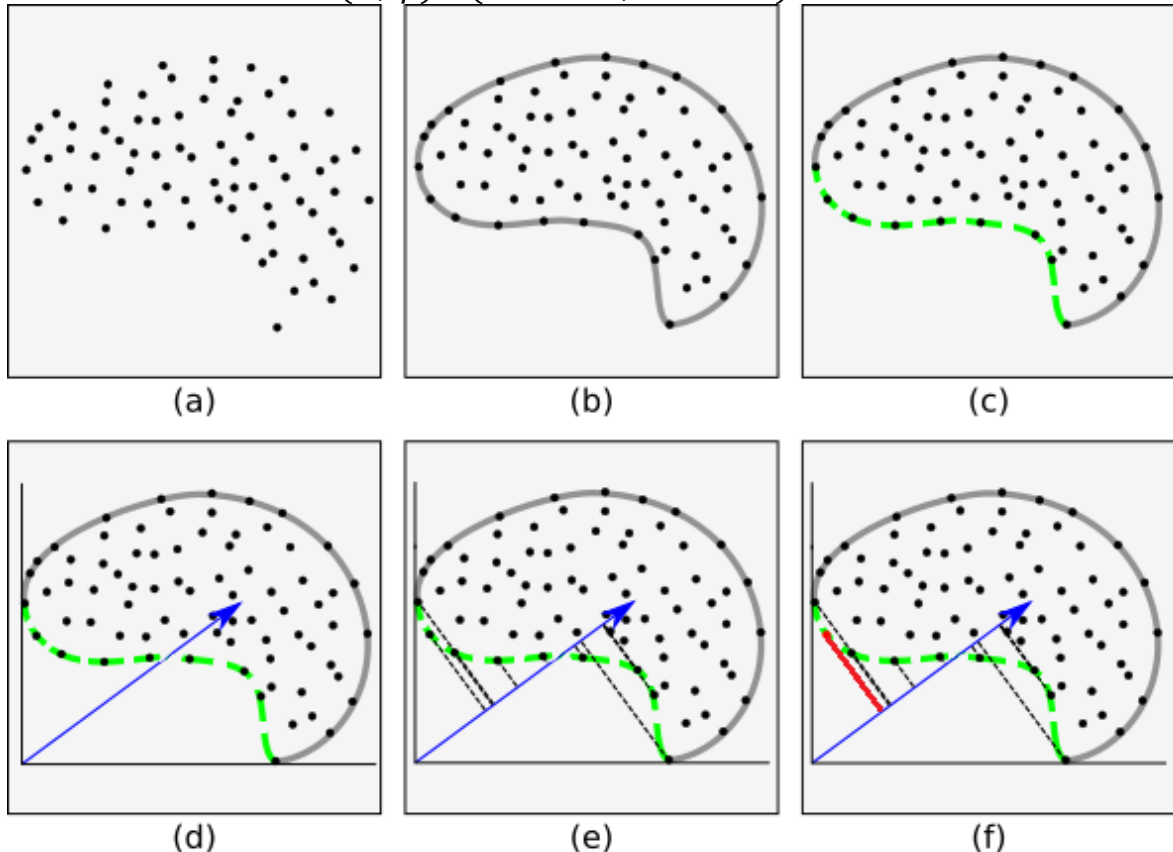


Рисунок 1 – Иллюстрация этапов предлагаемого метода МТМС

На рисунке 1 показана последовательность шагов МТМС для получения оптимального решения для заданных критериев. МТМС дает оптимальные по Парето решения. Шаги метода состоят в следующем:

Шаг 1. Определяем векторы из оценки V (количество таких векторов $N_{criteria}$), которые находятся в пространстве заданных критериев.

Шаг 2. Из Парето решения выбираем наиболее близкие к оптимальным по одному критерию и дальше от оптимальных по другим критериям. Получаем оптимальные по Парето решения $\hat{V} \subset V$ – ближайший Парето-фронт к началу координат пространства критериев:

$$\hat{V} = ParetoFront(V), \hat{V} \in \mathbb{R}^{N_{opt} \times N_{criteria}} \quad (8)$$

где N_{opt} – количество оптимальных по Парето решений.

Шаг 3. Оптимальные решения $\hat{v} \in \hat{V}$ масштабируем по каждому критерию до интервала $[0; 1]$:

$$\hat{V}_{scaled} = \frac{\hat{V}_i - \hat{v}_{min}}{\hat{v}_{max} - \hat{v}_{min}}, \hat{v}_{max} \in \mathbb{R}^{N_{criteria}}, \hat{v}_{min} \in \mathbb{R}^{N_{criteria}}, i = 1..N_{opt} \quad (9)$$

где \hat{v}_{max} – вектор максимальных значений \hat{V} для каждого критерия, \hat{v}_{min} – вектор минимальных значений \hat{V} для каждого критерия.

Таким образом, оптимальное решение – это решение, ближайшее к началу координат, и если какое-либо решение \hat{v} является началом координат, то оно оптимально для любого ϕ .

Шаг 4. Определяем вектор критериев ϕ .

Шаг 5. Определяем гиперпараметр в пространстве критериев наиболее близкий к данному критерию. Для этого спроецируем векторы из матрицы \hat{V}_{scaled} на вектор ϕ :

$$\hat{V}_{proj} = \frac{\hat{V}_{scaled}^T \phi}{\|\phi\|}, \hat{V}_{scaled} \in \mathbb{R}^{N_{opt}}. \quad (10)$$

6. Находим гиперпараметры θ , при которых минимум вектора \hat{V}_{proj} , что эквивалентно поиску минимальной взвешенной суммы значений вектора:

$$\theta = argmin_{\theta} \hat{V}_{proj}. \quad (11)$$

С МТМС проведены эксперименты по выбору оптимальных гиперпараметров с различными критериями.

На рисунке 2 показаны полученные результаты по выбору оптимальных гиперпараметров.

Разработан метод оптимизации вычислений нейронной сети, основное содержание которого представлено ниже.

1. Для поиска преобразования модели нейронной сети выполняется минимизация следующей оптимизационной функции:

$$L(U, F) = \frac{S(U, F) - S(U_0, F)}{S(U_0, F)} + \frac{T(U, F) - T(U_0, F)}{T(U_0, F)} + \frac{A(U, F) - A(U_0, F)}{A(U_0, F)} = \Delta_r S(U, F) + \Delta_r T(U, F) + \Delta_r A(U, F) \quad (12)$$

где F – функция преобразования входных данных в распределение вероятностей принадлежности к определенному классу объекту, U – функция изменения параметров F (такими параметрами являются весовые коэффициенты нейронной сети), U_0 – отсутствие изменения весовых коэффициентов, S – размер файла весовых коэффициентов нейронной сети, T – время обработки одного изображения нейронной сетью, A – точность нейронной сети на тестовой выборке.

2. Определяется функция изменения весовых коэффициентов U :

$$U = \operatorname{argmin}_{U \in \mathcal{U}} L(U, F) \quad (13)$$

где \mathcal{U} – множество функций изменения весовых коэффициентов нейронной сети.

Шаги предлагаемого метода оптимизации вычислений нейронной сети:

Шаг 1. Обучить нейронную сеть для заданной задачи.

Шаг 2. Выбрать множество функций преобразования весовых коэффициентов нейронной сети.

Шаг 3. Для каждой из этих функций произвести преобразование весовых коэффициентов нейронной сети.

Шаг 4. Для преобразованных нейронных сетей вычислить значений функции L по формуле (12).

Шаг 5. Выбрать функцию преобразований весовых коэффициентов, которое соответствует наименьшему значению L , то есть решить задачу (13).

Для выполнения преобразования весовых коэффициентов предлагается использовать квантование. Для квантования использовалась программа TensorFlow Lite.

Исследование по оптимизации выполнено для сети по нескольким видам квантования: квантование динамического диапазона (U_1), квантование с использованием репрезентативного набора данных (U_2), целочисленное квантование с использованием репрезентативного набора данных (U_3) и квантование во float16 (U_4). U_0 соответствует отсутствию квантованию (исходная модель без преобразований). В таблице 9 приведены результаты расчетов.

ϕ_0	ϕ_1	ϕ_2	ϕ_3	θ
0.5	0.5	0.5	0.5	base_lr=0.01, max_lr=0.01, cyclic_mode=triangular
0.0	0.5	0.5	0.5	base_lr=0.01, max_lr=0.01, cyclic_mode=triangular
1.0	0.5	0.5	0.5	base_lr=0.01, max_lr=0.01, cyclic_mode=triangular
0.5	0.0	0.5	0.5	base_lr=0.005, max_lr=0.0001, cyclic_mode=triangular
0.5	1.0	0.5	0.5	base_lr=0.01, max_lr=0.01, cyclic_mode=triangular
0.5	0.5	0.0	0.5	base_lr=0.01, max_lr=0.01, cyclic_mode=triangular
0.5	0.5	1.0	0.5	base_lr=0.01, max_lr=0.01, cyclic_mode=triangular
0.5	0.5	0.5	0.0	base_lr=0.0001, max_lr=0.005, cyclic_mode=triangular2

Рисунок 2 – Оптимальные гиперпараметры с соответствующими коэффициентами значимости критериев выбора

Таблица 9. Характеристики нейронной сети после каждого типа квантования

	Точность на тестовой выборке, %	Точность на обучающей выборке, %	Время/изображение, мс	Размер файла, КБ	L
Исходная модель (U_0)	99,18	99,92	23	3200	0,00
Динамический диапазон (U_1)	90,26	92,12	47	834	0,21
Использование набора данных (U_2)	99,18	99,92	483	999	19,31
Использование набора данных (целочисленный вход) (U_3)	99,08	99,94	486	1000	19,44
Использование float16 (U_4)	99,18	99,92	23	1612	-0,50

Из табл. 9 видно, что лучшим квантованием обладает квантование float16 (U_4), при котором достигается наименьшее значение (-0,50) оптимизируемой функции L .

В **четвёртой главе** описывается реализация прототипа системы автоматического распознавания на базе автомата по сортировке бытовых отходов «Сортомат». В прототипе были реализованы предложенные методы МТМС и оптимизации вычислений нейронной сети на основе квантования float16.

Для проведения эксперимента была использована нейронная сеть MobileNet, обученная классифицировать пластиковые (PET, HDPE) и алюминиевые банки. Для обучения нейронной сети создана выборка из 20000 изображений бутылок и банок. Для моделирования нескольких подсистем распознавания АСУТП выборка изображений разделена на три подвыборки, каждая из которых сделана в различные промежутки времени с разной загрязненностью крыльчатки. Таким образом, эти подвыборки имеют различное распределение данных.

В качестве критериев оптимизаций использовалась максимальная достигаемая точность и скорость сходимости обучения в соотношении 1:1, которые использовались в МТМС. Эти критерии дают гиперпараметры, оптимальные как с точки зрения точности, так и с точки зрения времени обучения, с одинаковой значимостью обоих критериев.

Для обученной нейронной сети проведена оптимизация с помощью разработанного метода оптимизации вычислений математической модели, в которой использовалось квантование float16.

В таблице 10 представлены результаты проведения эксперимента на трех задачах с разработанным методом оптимизации гиперпараметров.

Таблица 10. Результаты проведения эксперимента с разработанным методом оптимизации гиперпараметров на трех задачах классификации

	Максимальная точность на проверочной выборке, %	Номер эпохи, на которой нейронная сеть достигла сходимости	Размер исходного файла модели, КБ	Размер файла модели после квантования, КБ	На сколько уменьшился файл модели, %
Задача 1	96,61	15	12527,05	6278,27	49,88
Задача 2	96,69	15	12527,29	6278,62	49,88
Задача 3	93,08	15	12527,05	6278,27	49,88

Нейронная сеть при использовании ранее выбранных параметров аугментации и гиперпараметров обучения на всех трех задачах достигает достаточно высокой точности классификации (не менее 93 %), а также на всех трех задачах модель уменьшилась примерно на 50%.

Внедрение обученной с помощью разработанного прототипа системы самообучения и метода оптимизации гиперпараметров проводилось в «Сортомат», на который получен патент на полезную модель.

На рисунке 3 показан внешний вид «Сортомата». Слева расположен дисплей, на котором отображается информация, что предмет принят или нет, справа вверху отсек приема, внизу – отсек возврата, если предмет распознан как прочий мусор или как предмет, не пригодный для приема.

На рисунке 4 показано сопоставление результатов испытаний «Сортомата» и проведенного моделирования. Сопоставление производится по двум критериям: точность и номер эпохи.

На рисунке 5 показаны сопоставление гиперпараметров без их выбора и выбранные с помощью МТМС гиперпараметры. Сопоставление производится по двух критериям: точность и номер эпохи обучения нейронной сети, на которой достигается максимальная точность.



Рисунок 3 – Внешний вид «Сортомата»

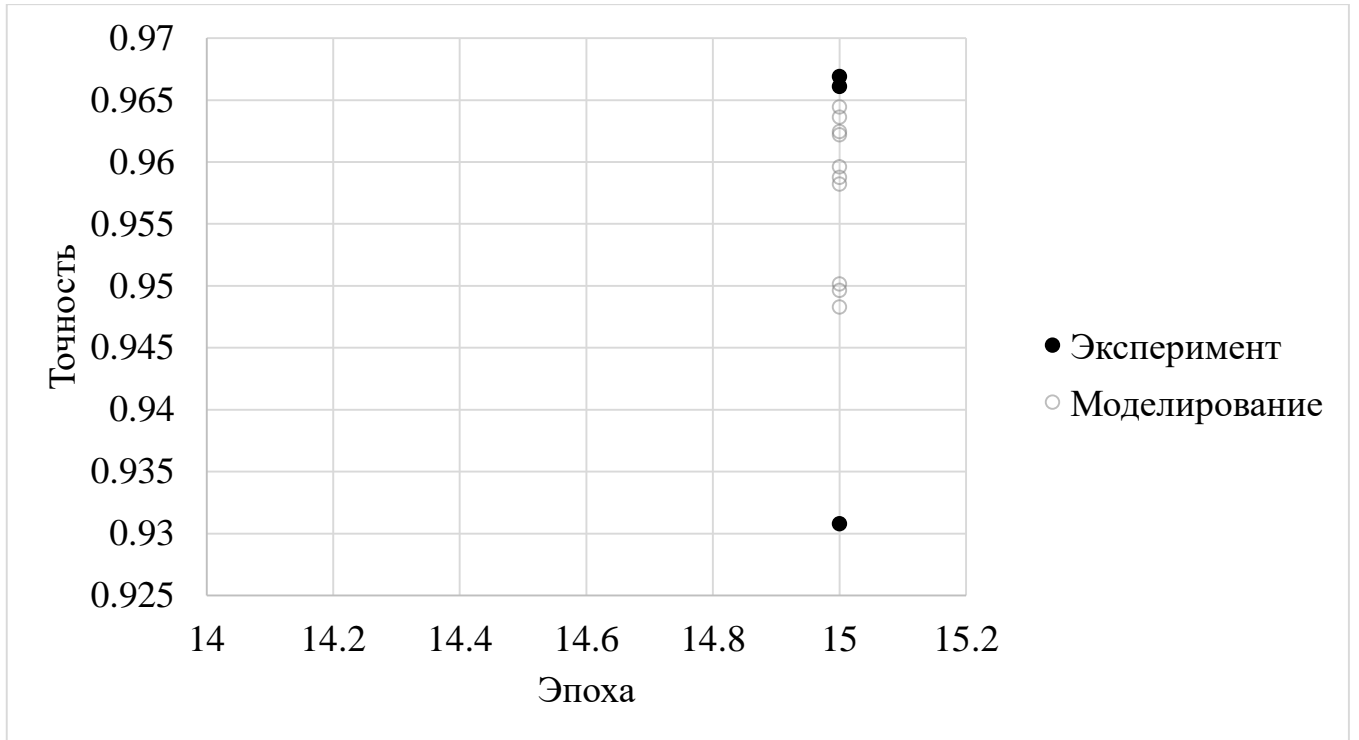


Рисунок 4 – Сопоставление экспериментальных результатов и результатов моделирования

Для сравнения точностей нейронной сети, полученной без выбора гиперпараметров и с выбранными гиперпараметрами, проведем множественную проверку гипотез, применив поправку Бонферрони (12):

$$\left\{ \begin{array}{l} X_1^n = (X_{11}, \dots, X_{1n}), X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \\ X_2^n = (X_{21}, \dots, X_{2n}), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2), \\ H_0: \mu_1 = \mu_2, \\ H_1: \mu_1 < \mu_2, \\ D_{1i} = X_{1i} - X_{2i}, \\ S_{d1} = \sqrt{\frac{1}{n-1} \left(\sum_i D_{1i}^2 - \frac{(\sum_i D_{1i})^2}{n} \right)}, \\ t_1 = \frac{\mathbb{E}X_1^n - \mathbb{E}X_2^n}{\frac{S_{d1}}{\sqrt{n}}}, \\ T(X_1^n, X_2^n) \sim St(n-1), \end{array} \right. \quad (12)$$

где X_1^n – точности, полученные на тестовых выборках для обученной нейронной сети без выбора гиперпараметров, X_2^n – точности, полученные для нейронной сети с выбором гиперпараметров, D_{1i} – попарные разности точностей из X_1^n и X_2^n , S_{d1} – стандартное отклонение разностей для выборок точностей.

После проведения вычислений установлено, что точность, достигаемая с помощью выбранных гиперпараметров, выше на 3%, чем без выбора гиперпараметров. Вместе с применением разработанного метода увеличения точности на основе аугментации увеличение точности достигает 23%.

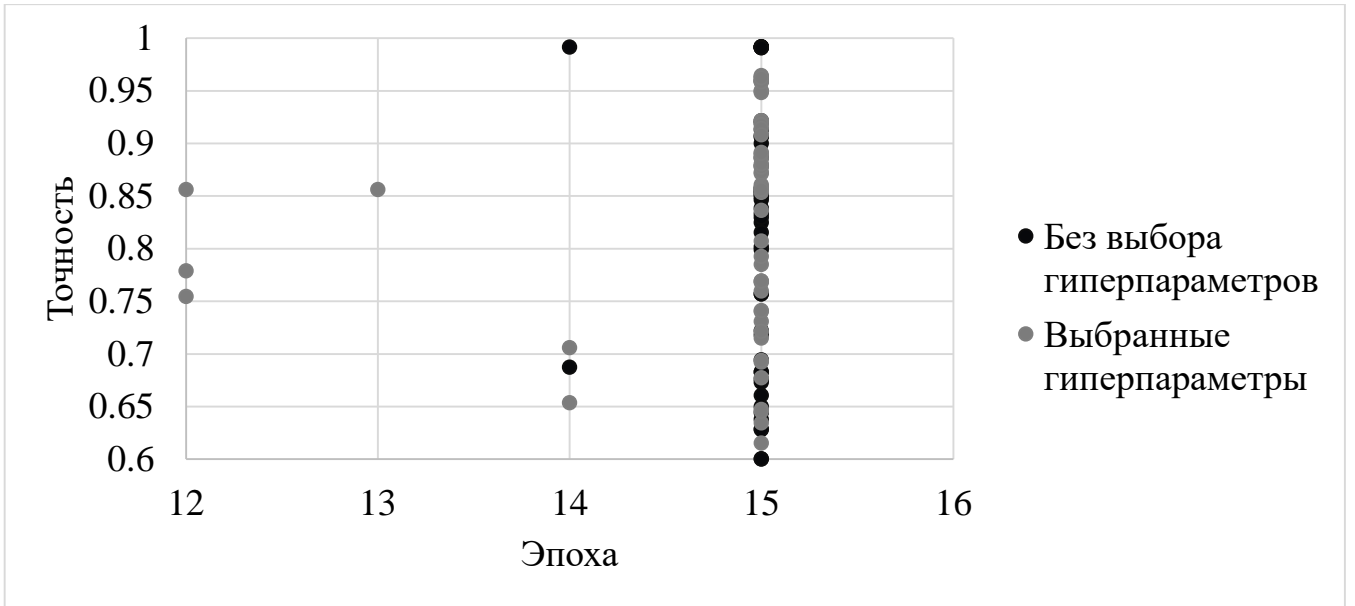


Рисунок 5 – Сопоставление результатов обучения нейронной сети для гиперпараметров по умолчанию и выбранных гиперпараметров

Для сравнения точностей нейронной сети, полученных в ходе экспериментов в «Сортомате» и моделировании, проведем множественную проверку гипотез (13):

$$\left\{ \begin{array}{l}
 X_5^{n_1} = (X_{51}, \dots, X_{5n_1}), X_5 \sim \mathcal{N}(\mu_5, \sigma_5^2), \\
 X_6^{n_2} = (X_{61}, \dots, X_{6n_2}), X_6 \sim \mathcal{N}(\mu_6, \sigma_6^2), \\
 H_0: \mu_5 = \mu_6, \\
 H_1: \mu_5 \neq \mu_6, \\
 t_3 = \frac{\mathbb{E}X_5^{n_1} - \mathbb{E}X_6^{n_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \\
 v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}, \\
 T(X_5^{n_1}, X_6^{n_2}) \sim St(v),
 \end{array} \right. \quad (13)$$

где $X_5^{n_1}$ – точности, полученные на тестовых выборках при моделировании, $X_6^{n_2}$ – точности, полученные для нейронной сети при проведении эксперимента, v – число степеней свободы для двух независимых выборок.

После проведения вычислений установлено, что точности нейронной сети при моделировании и эксперименте, сопоставимы, расхождение не более 1% при сравнении математических ожиданий точностей.

После внедрения разработанных методов получены следующие результаты:

1. Применение разработанного метода гиперпараметрической оптимизации увеличивает точность распознавания бытовых отходов на 3%, в совокупности с другими разработанными методами увеличение точности распознавания достигает 23%.

2. Полученные результаты подтверждаются актами внедрения в «Сортомат» и в учебный процесс.

3. Оригинальность разработанной программы подтверждает полученное свидетельство.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Основным результатом диссертационной работы является решение научной задачи, заключающейся в разработке научно-методического аппарата автоматической классификации изображений с помощью сверточных нейронных сетей, позволяющего уменьшить время обучения нейронной сети без потери точности. В ходе работы:

1. Проведён анализ принципов функционирования существующих систем автоматизированной сортировки бытовых отходов для выявления возможности оптимизации использования вычислительных ресурсов.

2. Предложен метод построения критериев оптимизации гиперпараметров для распознавания бытовых отходов, метод гиперпараметрической оптимизации обучения сверточной нейронной сети с заданными критериями и метод оптимизации вычислений математической модели.

3. Предложен метод обучения сверточной нейронной сети.

4. Предложены алгоритмы самообучения в автоматизированной системе сортировки бытовых отходов.

5. Внедрение разработанного программного комплекса в «Сортомат» показало, что достигается высокая точность (не менее 93%) классификации изображений и при этом размер файла нейро-сетевой модели уменьшается примерно на 50%.

ОСНОВНЫЕ ПОЛОЖЕНИЯ В ДИССЕРТАЦИИ ОТРАЖЕНЫ В СЛЕДУЮЩИХ ПУБЛИКАЦИЯХ

Статьи, опубликованные в изданиях ВАК РФ:

1. Ахметзянов, К. Р. Оптимизация вычислений нейронной сети / К. Р. Ахметзянов, А. И. Тур, А. Н. Кокоулин, А. А. Южаков // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2020. – № 36. – С. 117-130.

2. Ахметзянов, К. Р. Разработка и исследование модели иерархической системы распознавания объектов для оценки влияния параметров изображений на быстродействие распознавания / А. И. Тур, А. Н. Кокоулин, К. Р. Ахметзянов [и др.] // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2020. – № 35. – С. 87-101.

3. Ахметзянов, К. Р. Вопросы применения иерархических систем распознавания в системах видеонаблюдения / А. И. Тур, А. Н. Кокоулин, К. Р. Ахметзянов, А. А. Южаков // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2020. – № 34. – С. 75-89.

4. Ахметзянов, К. Р. Разработка нейронной сети на основе знаний о воздействиях окружающей среды / К. Р. Ахметзянов, А. А. Южаков // *Нейрокомпьютеры: разработка, применение.* – 2019. – Т. 21. – № 3. – С. 5-13.

5. Ахметзянов, К. Р. Оптимизация архитектуры нейронной сети для сортировки групп мусора / К. Р. Ахметзянов, А. А. Южаков // *Известия СПбГЭТУ ЛЭТИ.* – 2019. – № 7. – С. 62-70.

6. Ахметзянов, К. Р. Увеличение точности сверточной нейронной сети за счет возрастания количества данных / К. Р. Ахметзянов, А. А. Южаков // *Нейрокомпьютеры: разработка, применение.* – 2018. – № 7. – С. 14-19. – DOI 10.18127/j19998554-201807-03.

7. Ахметзянов, К. Р. Сравнение сверточных нейронных сетей для задач сортировки мусорных отходов / К. Р. Ахметзянов, А. А. Южаков // *Известия СПбГЭТУ ЛЭТИ.* – 2018. – № 6. – С. 27-32.

Статьи, опубликованные в изданиях, индексируемых в базе Scopus:

8. Akhmetzyanov, K. R. Neural Network Development Based on Knowledge about Environmental Influence / K. R. Akhmetzyanov, A. A. Yuzhakov, A. N. Kokoulin // *Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering* – 2020. – P. 218-221.

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

10. Akhmetzyanov, K. R. Image Processing by Using Convolutional Neural Networks for Waste Sorting / K. R. Akhmetzyanov, A. A. Yuzhakov, A. N. Kokoulin, D. M. Filatov // *International Conference on Soft Computing and Measurements* – 2018. – Vol. 1. – P. 654-657.

Патент и свидетельство о регистрации программ для ЭВМ, полученные по тематике диссертационной работы:

11. Ахметзянов, К.Р. Патент на полезную модель № 188755 U1 Российская Федерация, МПК G07F 11/00. Автомат по приёму тары / А. И. Тур, А. Н. Кокоулин, К. Р. Ахметзянов, А. А. Южаков. – М.: Роспатент, 2019.

12. Ахметзянов, К. Р. Свидетельство о государственной регистрации программы для ЭВМ № 2019664159 Российская Федерация. Программа бинокулярного зрения с учетом расстояния до объекта / К. Р. Ахметзянов, А. И. Тур, А. Н. Кокоулин, А. А. Южаков. – М.: Роспатент, 2019.

Статьи, опубликованные в других изданиях:

13. Ахметзянов, К. Р. Моделирование системы иерархического распознавания для предварительного расчёта характеристик процесса / А. И. Тур, К. Р. Ахметзянов, А. Н. Кокоулин // *Интеллектуальные системы в науке и технике. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века* – 2020. – С. 228-233.

14. Ахметзянов, К. Р. Оптимизация архитектуры нейронной сети для сортировки мусорных отходов / К. Р. Ахметзянов, А. А. Южаков // *Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века* – 2019. – С. 82-83.