

Тип статьи: научная
УДК 631.365.34
DOI: 10.35887/2305-2538-2024-2-98-105

ВЫБОР МОДЕЛИ И АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ ДЛЯ СИСТЕМЫ ТЕХНИЧЕСКОГО ЗРЕНИЯ ДЕТЕКЦИИ КОРНЕПЛОДОВ САХАРНОЙ СВЕКЛЫ

Данила Александрович Николукин¹, Владимир Евгеньевич Петерс²,
Михаил Александрович Попов³, Андрей Владиленович Крищенко⁴

^{1,2} Тамбовский государственный технический университет, г. Тамбов, Российская Федерация
^{1,2,3,4} Всероссийский научно-исследовательский институт использования техники и
нефтепродуктов в сельском хозяйстве, г. Тамбов, Российская Федерация

¹JeyViJey@mail.ru, ¹<https://orcid.org/0009-0008-4860-677X>

²<https://orcid.org/0009-0000-1066-6752>

Автор ответственный за переписку: Данила Алексеевич Николукин, JeyViJey@mail.ru
Corresponding author: Danila Nikoluykin, JeyViJey@mail.ru

Реферат. Среди методов долгосрочного хранения корнеплодов сахарной свёклы рассматривается использование системы технического зрения, направленное на обеспечение большей сохранности сырья в период его уборки и хранения путём анализа прибывающего на заводы сырья. В работе рассмотрен процесс распределения грузовых автомашин на приёмном пункте сахарных заводов человеком. Целью статьи является разработка метода детекции корнеплодов сахарной свёклы на изображении поверхности насыпи грузовой автомашины на приёмном пункте сахарных заводов. Методом распознавания сырья выбрана нейронная сеть в связи с чрезвычайно разнообразными получаемыми изображениями, с чем не справится распознавание с помощью порогового преобразования. Рассмотрены алгоритмы (такие варианты, как SSD, YOLO, R-CNN) и архитектуры (такие варианты, как AlexNet, VGGNet, GoogleNet, ResNet) различных нейронных сетей. Выявлена наиболее подходящая под необходимые задачи нейронная сеть. Определён оптимальный размер базы данных для обучения. В результате обучения и тестирования нейронной сети, выявлена необходимость в создании дополнительных фильтров. Созданы и описаны фильтры, уточняющие работы нейронной сети: отсеивающие “слипшиеся” объекты, объекты с низкой принадлежностью классу, объекты с невозможной площадью (чересчур маленькой или большой), а также с соотношением сторон менее 0,2 или более 5. По результатам выполненной работы была обучена нейронная сеть, способная находить на поступающих изображениях достаточное количество корнеплодов сахарной свёклы и удовлетворяющая требованиям качественного функционирования системы технического зрения.

Ключевые слова: сахарная свекла, хранение, размер корнеплода сахарной свеклы, машинное зрение, кагат.

CHOOSING A NEURAL NETWORK MODEL AND ARCHITECTURE FOR A TECHNICAL VISION SYSTEM FOR DETECTING SUGAR BEET ROOT CROPS

Danila Nikoluykin¹, Vladimir Peters², Mikhail Popov³, Andrey Kryshchenko⁴

^{1,2} Tambov State Technical University, Tambov, Russia

^{1,2,3,4} Scientific Research Institute for the Use of Machinery and Petroleum Products in
Agriculture, Tambov, Russia

¹JeyViJey@mail.ru, ¹<https://orcid.org/0009-0008-4860-677X>

²<https://orcid.org/0009-0000-1066-6752>

Abstract. The use of a technical vision system, aimed at ensuring greater safety of sugar beet roots during harvesting and storage by analyzing raw materials arriving at factories, is considered among the methods of long-term storage. The process of distribution of trucks at the receiving point of sugar

factories is considered by humans in the work. The purpose of the work is to develop a method for detecting sugar beet roots on an image of the surface of a truck embankment at the receiving point of sugar factories. A neural network was chosen as the raw material recognition method due to the extremely diverse nature of the resulting images, which recognition using a threshold transformation cannot cope with. Algorithms (options such as SSD, YOLO, R-CNN) and architectures (options such as AlexNet, VGGNet, GoogleNet, ResNet) of various neural networks are considered. The most suitable neural network for the required tasks has been identified. The optimal size of the database for training has been determined. As a result of training and testing the neural network, the need to create additional filters was identified. Filters have been created and described that clarify the work of the neural network: filtering out "stuck together" objects, objects with low class membership, objects with an impossible area (too small or large), as well as with an aspect ratio of less than 0.2 or more than 5. A neural network capable to find a sufficient number of sugar beet roots in incoming images and to meet the requirements for high-quality functioning of the technical vision system, was trained based on the results of the work performed.

Keywords: *sugar beet, storage, sugar beet root crop size, machine vision, kagat.*

Для цитирования: Николюкин Д.А., Петерс В.Е., Попов М.А., Крищенко А.В. Выбор модели и архитектуры нейронной сети для системы технического зрения детекции корнеплодов сахарной свёклы // Наука в Центральной России Science in the Central Russia. 2024. Т. 68, № 2. С. 98-105. <https://doi.org/10.35887/2305-2538-2024-2-98-105>.

For citation: Nikolyukin D., Peters V., Popov M., Kryshchenko A. Choosing a neural network model and architecture for a technical vision system for detecting sugar beet root crops. *Nauka v central'noj Rossii* = Science in the Central Russia: 2024; 68(2): 98-105. (In Russ.) <https://doi.org/10.35887/2305-2538-2024-2-98-105>.

Введение. В свеклосахарной отрасли Российской Федерации актуальной является задача разработки метода сортировки при формировании кагатов сахарной свеклы для долгосрочного хранения сроком более двух месяцев. В настоящее время сортирует партии сахарной свеклы человек и делает это с погрешностью из-за человеческих факторов. Точность сортировки человеком по оценочным данным составляет 50-60%. Применение системы технического зрения дает возможность идентифицировать партии сахарной свеклы на партии с крупным сырьем, и на партии с мелким сырьем и ответить на вопрос: какие партии сырья отправлять сразу на переработку, а какие лучше сначала отправить на хранения для снижения рисков потери массы и качества сахарной свеклы.

Рассмотрим процесс распределения сахарной свёклы на заводе ООО «Промсахар» Курской области. Сырьё поступает на свёклоперерабатывающий завод с полей в большегрузных автомашинах. На территории завода автомашина доставляет сырьё на приёмный пункт, где заводской персонал производит визуальную оценку загрязнённости, засоренности, крупности сырья. После анализа, автомашина отправляется на выгрузку сырья либо в оперативные кагаты, либо в кагаты длительного вентилируемого хранения (ДВХ). Процесс распределения сырья представлен на рисунке 1.

Оперативные кагаты используются для своевременной сырьевой подпитки завода. ДВХ кагаты используются в качестве резерва и состоят преимущественно из крупной сахарной свёклы. Попадание в ДВХ кагаты партий мелкой сахарной свёклы приводит к значительному сокращению хранения (с полутора месяцев до двух недель). Это связано с тем, что скопления мелкой сахарной свёклы в ДВХ кагатах образуют участки повышенной плотности, которые плохо вентилируются. В связи с этим, эти участки создают зоны самосогревания, что приводит к повышению влажности этих зон и кагата в целом. Использование системы технического зрения исключит влияние человеческого фактора (усталость, невнимательность, торопливость) на анализ сырья, улучшит процесс распределения сырья по типам кагатов за счёт применения вычислительной техники и нейронных сетей.



Рисунок 1 – Процесс распределения поступающего на завод сырья

Материалы и методы. Так как завод может работать с самого утра и до самой ночи, наличие освещения даёт системе технического зрения необходимый уровень освещения в любое время суток и в любую погоду. Таким образом, исключается возможность получить полностью тёмное изображение из-за нехватки света изображения. Примеры насыпей представлены на рисунке 2.

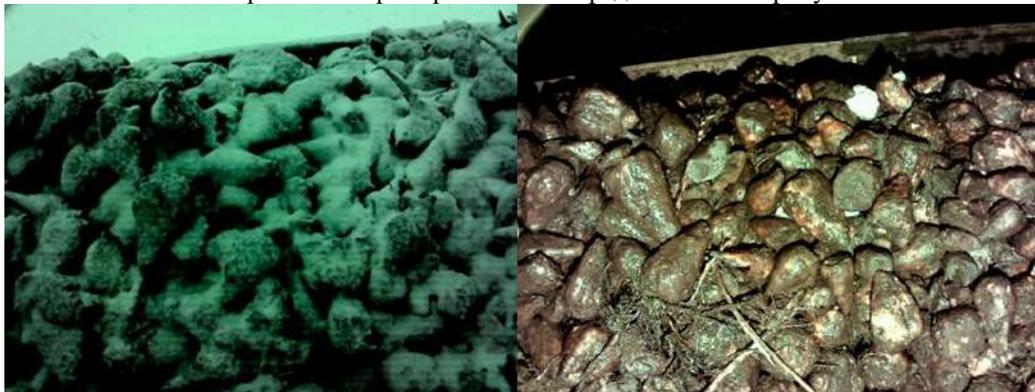


Рисунок 2 – Насыпи сахарной свеклы

В связи с чрезвычайно разнообразными погодными и временными условиями, а именно с утра до ночи, туман, снег, дождь, яркое солнце, облачная погода, использовать для детекции отдельных корнеплодов сахарной свеклы пороговое преобразование по цвету нецелесообразно и даст большую погрешность, поэтому лучше использовать нейронные сети.

Для обнаружения объектов (object detection) на изображениях широко используются алгоритмы обнаружения (детекторы), такие как SSD, YOLO, R-CNN [2].

Алгоритм R-CNN имеет точность обнаружения объекта – 58,5. Основной недостаток: медленная скорость обработки изображений. Распространенные версии: Fast R-CNN и Faster R-CNN. Принцип работы R-CNN: на изображении выделяются регионы (в конечном варианте получаются около 2000 регионов), которые предположительно содержат необходимый объект; регионы преобразуются к размеру входного слоя нейронной сети; фиксируются характеристики с

предпоследнего полносвязного слоя сети; характеристики классифицируются (SVM), и уточняется положение каждого региона на изображении (регрессия) [3].

Алгоритм YOLO имеет точность обнаружения – 63,4. Исходное изображение делится на ячейки. Если центр объекта содержится внутри определенной ячейки, то данную ячейку принимают за «главную», отражающую параметры местонахождения объекта. Каждая ячейка воспроизводит несколько вариантов нахождения ограничивающей рамки на изображении для одного и того же объекта. Данные варианты имеют определенные характеристики – координаты центра рамки, ширину и высоту рамки, степень «содержания объекта» в рамке. Последний слой используемой сети «принимает» конечное решение о количестве предсказываемых рамок ячеек и классификации объекта. Недостатки: плохое качество распознавание объекта сложной формы, группы небольших объектов [4].

Алгоритм SSD имеет точность обнаружения – 72,1. Изображения последовательно передаются на слои сверточной сети, уменьшающиеся в размерах. Выходы слоев сверточной сети каждой размерности учувствуют в детектировании, тем самым, формирую «пирамиду» изображения. Достоинства алгоритма: обнаружение объектов различных масштабов, более быстрая работа и точность обнаружения. SSD предсказывает смещение ограничивающих рамок. Рамки на разных уровнях масштабируются, размерность выходного слоя отвечает за объекты определенного масштаба. Большие объекты могут быть обнаружены на высоком уровне, маленькие – на низком [5].

Ориентируясь на вышеизложенные факты, для реализации системы технического зрения было решено использовать модель нейронной сети, предполагающую алгоритм SSD для обнаружения сложных объектов на изображениях. Теперь нужно выбрать архитектуру нейронной сети.

Архитектура сверточных сетей очень обширна. В качестве примеров, можно рассмотреть такие варианты, как AlexNet, VGGNet, GoogleNet, ResNet.

AlexNet – это одна из первых архитектур глубоких сверточных сетей. Она состоит из пяти сверточных слоев, между которыми располагаются pooling-слои и слои нормализации, и три полносвязных слоя в конце [5]. Недостатки архитектуры: AlexNet, в основном, используется в качестве основы для других новых архитектур и обладает меньшей точностью.

Архитектура VGGNet, по точности обнаружения объектов, превзошла AlexNet в 2 раза. Здесь используются большое число слоев с фильтрами меньших размеров (рисунок 18) [6].

Все ядра «conv» имеют размер 3x3, «maxpool» – 2x2 с шагом два. В этом заключается главная особенность данной архитектуры – использование ядер свертки небольших размеров, в отличие от AlexNet, что уменьшает количество обучаемых переменных. Благодаря этому, повышается устойчивость к переоснащению и обучение воспроизводится более быстро. Недостатки архитектуры: большой расход памяти.

GoogleNet – архитектура с 22 слоями. Обладает большей вычислительной способностью. GoogleNet состоит из модулей Inception, расположенных друг за другом. Имеет 5 миллионов параметров, также не содержит в себе полносвязные слои. В начале сети располагается подсеть Stem Network, в состав которой входят три сверточных слоя с двумя слоями пулинга. Модуль Inception параллельно применяет несколько фильтров на изображение, результат воздействия которых преобразуется в выходной сигнал, который перенаправляется на следующий слой [7]. Архитектура также предполагает наличие слоев с фильтром 1x1 («узкие места»), которые необходимы для уменьшения глубины изображения. GoogleNet отличается заметной эффективностью и точностью, равной 93,3 %. Однако качество данной сети существенно отличается от функционала более поздней усовершенствованной сети ResNet, с точностью 96,5 % [8].

ResNet, в отличие от GoogleNet, предполагает более 150 слоев, что делает ее в 2 раза эффективнее.

Таким образом, анализ архитектур современных нейронных сетей для распознавания объектов показал, что архитектура ResNet имеет наибольшую эффективность и была выбрана для решения поставленной задачи распознавания свеклы в насыпи кузова грузового автомобиля.

Модель выбранной для решения задачи нейронной сети носит название «SSD ResNet50», что расшифровывается как «Single-shot detector Deep residual network». Многие современные

исследования показывают, что данная модель наиболее эффективна для обнаружения объектов на изображениях при наличии небольших наборов данных. Преимуществом ResNet также могут служить «смягчение» эффекта проблемы исчезающих градиентов и улучшение точности (при свёртке глубокой сети увеличивается глубина сети и точность ухудшается).

На основе описанной выше простой сети добавлено быстрое соединение, которое превращает сеть в ее остаточную версию. Идентификационные быстрые соединения $F(x\{W\} + x)$ могут использоваться непосредственно, когда вход и выход имеют одинаковые размерности. Когда размерности увеличиваются, он рассматривает два варианта: Быстрое соединение выполняет сопоставление идентификаторов с дополнительными нулями, добавленными для увеличения размерности. Эта опция не вводит никаких дополнительных параметров. Проекция быстрого соединения в $F(x\{W\} + x)$ используется для сопоставления размерностей (выполнено с помощью 1×1 свертки). Для любой из опций, если быстрые соединения идут по картам объектов двух размерностей, они выполняются с шагом 2.

В интеллектуальной системе технического зрения принято использовать 50-слойную ResNet. В этой конфигурации сети каждый 3-слойный блок заменяется в 34-слойной сети этим 3-слойным узким местом, и в результате получается 50-слойная ResNet (рисунок 3). Они используют вариант 2 для увеличения размерностей. Эта модель имеет 3,8 миллиарда FLOPs.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Рисунок 3 – Сравнение конфигураций ResNet

Результаты и их обсуждение. Для обучения нейронной сети была собрана база изображений сахарной свеклы в кузове грузовых автомобилей. Количество снимков составило 258. Принимая во внимание вышеизложенное описание свойств сети ResNet, количество изображений является достаточным для обучения и для тестирования сети. На изображениях с использованием программы labeling вручную были размечены клубни сахарной свеклы. Результатом этого шага стали изображения и соответствующие им xml-файлы с размеченными данными о координатах каждого клубня.

Для процесса обучения изображения были разделены на обучающую и тестирующую выборки в соотношении 2/3 (172 изображений) и 1/3 (86 изображений) от общего количества снимков соответственно. Результаты обучения представлены на рисунке 4. Если брать большую или меньшую базу данных, результат менялся в обоих случаях в худшую сторону, поэтому принято решение оставить количество изображений таким.

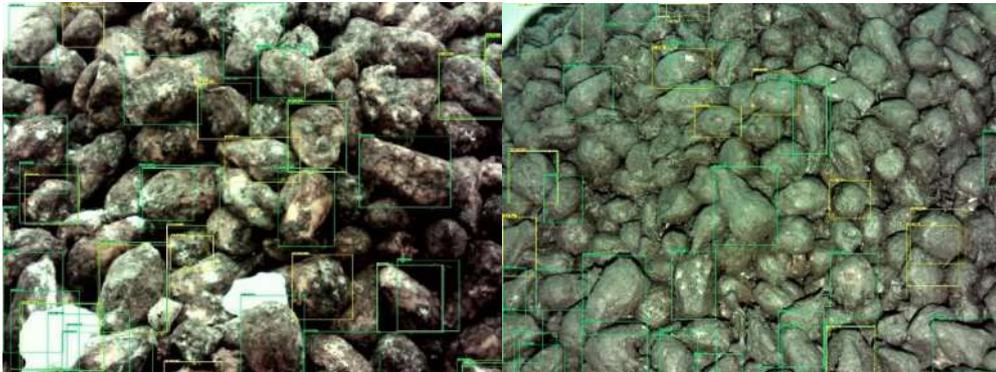


Рисунок 4 – обработанные изображения нейронной сетью

В результате обработки обученной нейронной сетью изображений получается, что некоторые корнеплоды классифицируются двумя классами (свёкла сверху и свёкла сбоку). Это сильно мешает адекватной работе системы технического зрения. Помимо этого, происходит детекция сразу двух, трёх или большего количества объектов как один. Для улучшения распознавания корнеплодов нейронной сетью необходимы дополнительные фильтры.

Фильтры работают по следующему алгоритму: удаление из массивов элементов, принадлежность к классу которых меньше 5%, определение координат двух противоположных вершин прямоугольника, ограничивающего область, содержащую корнеплод, отсеивание элементов, чьи координаты совпадают, отсеивание элементов с площадью от 500 тысяч пикселей или соотношением сторон больше 5 или меньше 0,2. Чтобы исключить наложение двух классов на один объект происходит отсеивание одного из них путём сравнения вероятности принадлежности к классу: меньшая вероятность удаляется вместе со всеми данными. Оставшийся массив считается правильным, так как больше вероятность принадлежности. Чтобы исключить идентификацию двух и более объектов за один происходит отсеивание объектов со слишком большой площадью путём сравнения площади в пикселях с фиксированным значением в 500 тысяч пикселей. Выше этого значения корнеплод даже очень крупного размера быть не может. Помимо этого, сравнивается и соотношение длины к ширине. Если соотношение больше 5 или меньше 0,2, значит этот объект отсеивается.

Применив эти фильтры, качество детекции заметно улучшилось. Примеры изображений, обработанных нейронной сетью с использованием фильтров представлены на рисунках 5 и 6.



Рисунок 5 – Обработанное нейронной сетью и фильтрами изображение

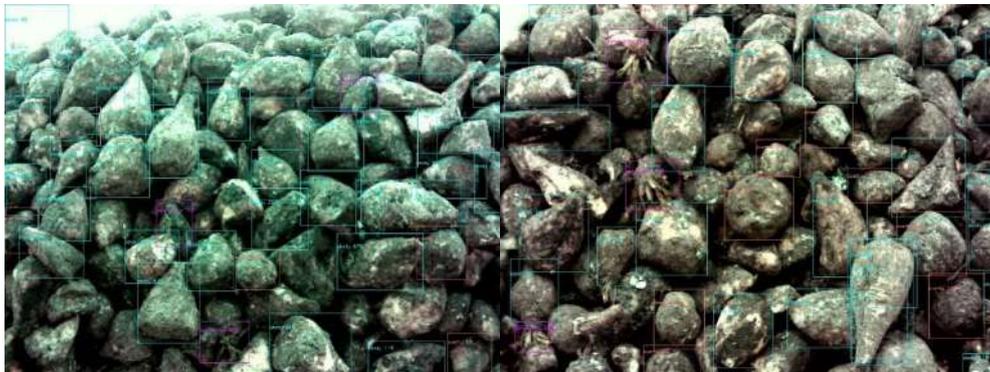


Рисунок 6 – Обработанное нейронной сетью и фильтрами изображение

По результатам использования фильтров для нейронной сети, можно сказать, что результат является удовлетворительным по сравнению с существующим процессом зрительного анализа человеком.

Заключение. В результате проведённой работы была доказана необходимость применения нейронных сетей в системе детекции корнеплодов сахарной свёклы на свеклопункте. Используя нейронные сети, можно снизить риски потери качества сахарной свёклы в период хранения.

Список источников

1. Завражных А.И. Обоснование использования машинного зрения для сортировки сахарной свёклы при хранении в кагатах / А.И. Завражных, С.М. Кольцов // Достижения науки и техники АПК. - 2022. - Т.36, №12. - С. 59 – 62.
2. Hobson Jane. Natural language processing in action / Jane Hobson, Cole Howard, Hannes Max Napke – Shelter Island, NY 11964: Manning Publications Co, 2019. – 513 с.
3. Yves Hilpisch. Artificial Intelligence in Finance. A Python-Based Guide / Hilpisch Yves – Sebastopol, CA 95472.: O’Reilly Media, 2020. – 477 с.
4. Eli Stevens. Deep Learning with PyTorch / Stevens Eli, Antiga Luca, Viehmann Thomas – Shelter Island, NY 11964: Manning Publications Co, 2020. – 522 с.
5. Paul Azunre. Transfer Learning for Natural Language Processing / Azunre Paul – Shelter Island, NY 11964: Manning Publications Co, 2021. – 266 с.
6. Oliver Durr. Probabilistic Deep Learning. With Python, Keras and TensorFlow Probability / Durr Oliver, Sick Beate, Murina Elvis – Shelter Island, NY 11964: Manning Publications Co, 2020. – 297 с.
7. Aurelien Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems / Geron Aurelien – Sebastopol, CA 95472.: O’Reilly Media, 2019. – 1065 с.
8. Rishal Hurbans. Artificial Intelligence Algorithms / Hurbans Rishal – Shelter Island, NY 11964: Manning Publications Co, 2020. – 393 с.

References

1. Zavrazhnov A.I. Justification of the use of machine vision for sorting sugar beet during storage in kagats / A.I. Zavrazhnov, S.M. Koltsov // Achievements of science and technology of the agro-industrial complex. - 2022. - vol.36, No.12. – pp. 59-62.
2. Hobson Jane. Natural language processing in action / Jane Hobson, Cole Howard, Hannes Max Napke – Shelter Island, NY 11964: Manning Publications Co, 2019. – 513 с.
3. Yves Hilpisch. Artificial Intelligence in Finance. A Python-Based Guide / Hilpisch Yves – Sebastopol, CA 95472.: O’Reilly Media, 2020. – 477 с.
4. Eli Stevens. Deep Learning with PyTorch / Stevens Eli, Antiga Luca, Viehmann Thomas – Shelter Island, NY 11964: Manning Publications Co, 2020. – 522 с.

5. Paul Azunre. Transfer Learning for Natural Language Processing / Azunre Paul – Shelter Island, NY 11964: Manning Publications Co, 2021. – 266 с.

6. Oliver Durr. Probabilistic Deep Learning. With Python, Keras and TensorFlow Probability / Durr Oliver, Sick Beate, Murina Elvis – Shelter Island, NY 11964: Manning Publications Co, 2020. – 297 с.

7. Aurelien Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems / Geron Aurelien – Sebastopol, CA 95472.: O'Reilly Media, 2019. – 1065 с.

8. Rishal Hurbans. Artificial Intelligence Algorithms / Hurbans Rishal – Shelter Island, NY 11964: Manning Publications Co, 2020. – 393 с.

Информация об авторах

Д.А. Николюкин – младший научный сотрудник; В.Е. Петерс – младший научный сотрудник; М.А. Попов – младший научный сотрудник; А.В. Крищенко – инженер.

Information about the authors

D. Nikol'yukin – junior researcher; V. Peters – junior researcher; M. Popov – junior researcher; A. Krischenko – engineer.

Вклад авторов все авторы сделали эквивалентный вклад в подготовку публикации.

Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: all authors have made an equivalent contribution to the preparation of the publication.

The authors declare no conflict of interest.

Поступила в редакцию (Received): 26.02.2024 Принята к публикации (Accepted): 12.04.2024