

УДК 004.032.26
DOI: 10.15827/0236-235X.131.464-470

Дата подачи статьи: 13.12.19
2020. Т. 33. № 3. С. 464–470

Проецирование технических объектов на изображении в метрическое пространство при помощи глубоких нейронных сетей для задачи детектирования

О.В. Толстель¹, к.т.н., доцент, tolstel.oleg@mail.ru

А.Е. Ширкин¹, аспирант, shirkin39@mail.ru

А.А. Калабин², д.ф.-м.н., профессор, akalabin@yandex.ru

¹ Балтийский федеральный университет им. И. Канта,
Институт физико-математических наук и информационных технологий,
г. Калининград, 236041, Россия

² Тверской государственный технический университет, г. Тверь, 170026, Россия

Представлен алгоритм векторизации изображения, содержащего технические объекты. В качестве таких объектов выступают машиностроительные детали, элементы крепежа, метизы. Под векторизацией подразумевается преобразование изображения в вектор, для которого евклидово расстояние имеет семантический смысл. Данный алгоритм создавался для усовершенствования системы оценки позиции объектов, где существует проблема непостоянного количества типов предметов для распознавания.

Предложен подход к формированию метрического пространства, где изображение, преобразованное в вектор по метрике l_2 , может сравниваться с изображением-эталоном, тем самым решается задача непостоянного количества классов. Для добавления нового класса достаточно внести в систему изображение-эталон в виде вектора и найти расстояние до него. Если оно меньше других изображений, то данный эталон будет типом объекта, представленным на вход системе. Данный подход реализуется при глубоких нейронных сетях, где последний слой убирается, а оставляется предпоследний слой, который представляет верхний уровень признаков, извлеченных из изображения. Такая нейронная сеть проходит процесс обучения с помощью функции потерь Triplet loss, научив нейронную сеть векторизовать изображение в метрическое пространство.

Программа, реализующая предложенный алгоритм, разработана на языке Python 3.6 с использованием интегрированной среды Jupyter Lab для операционной системы Ubuntu 18.04.

Приведены результаты эксперимента по использованию предложенного алгоритма для отнесения полученных изображений к тому или иному эталону. Для оценки качества алгоритма использованы метрики ранжирования для задач поиска, где оценивается только самый первый объект в списке ближайших объектов.

Разработанный алгоритм может применяться в системах технического зрения для роботов-манипуляторов, а в дальнейшем в системах управления захватом объектов этим роботом.

Ключевые слова: глубокие нейронные сети, метрическое пространство, распознавание образов, захват объектов, роботы-манипуляторы.

В современных задачах компьютерного зрения и робототехники очень часто применяются алгоритмы машинного обучения, в частности, глубокие нейронные сети. Это обусловлено тем, что алгоритмы машинного обучения достигают наилучших результатов в решении многих задач, особенно компьютерного зрения и механики роботов.

Одна из задач робототехники, применяемая для захвата, – обнаружение и классификация объектов на изображении. Как правило, она решается многоклассовым детектированием алгоритмами R-CNN [1], Fast-RCNN [2], Faster-RCNN [3], Mask-RCNN [4], YOLO v3 [5]. Алгоритмы [1–4] являются развитием семейства

алгоритмов детектирования, где с каждым из них в порядке возрастания улучшаются качество и функциональность. На данный момент наилучшим в этой линейке является Mask-RCNN [4], но его использование не всегда требуется, так как, помимо детектирования, он еще решает и задачу сегментирования, когда для целевого класса выделяется не только прямоугольная рамка, где он находится на изображении, но и четкий контур, внутри которого находится объект. YOLO v3 [5] является лучшим алгоритмом в отдельной ветке алгоритмов детектирования YOLO. Алгоритмы семейства YOLO используются в задачах, где нужен баланс скорости и качества работы на данных.

Качество и скорость алгоритмов тестируются на наборах данных Microsoft COCO [6] и ImageNet Classification Challenge (IGCC) [7], так как эти данные содержат более 1 000 различных классов с 10 миллионами изображений. Однако у существующих подходов есть проблема в фиксированном количестве классов. Когда необходимо распознать новый класс, построенная система распознавания не способна определить новый предмет.

С недавнего времени начали появляться подходы, называемые Single Shot Learning (обучение с первого примера). Они основаны на создании некоторого отображения из картинки в n -мерный вектор таким образом, что метрика расстояния между данными векторами имеет семантический смысл. На основе отображения строится псевдометрическое пространство, в котором похожие по содержанию изображения находятся близко друг к другу, а отличающиеся далеко. Использование данной концепции позволяет соотносить объект-кандидат на изображении со списком изображений-эталонов, которые не привязаны к системе и могут добавляться и удаляться. В настоящее время для реализации данного подхода используются такие функции потерь, как Triplet Loss [8], Proxy-NCA [9] и Margin Loss [10], являющиеся так называемыми функциями потерь ранжирования. В отличие от привычной перекрестной энтропии или среднеквадратичной ошибки, чья основная цель – предсказывать определенный класс, значение или набор значений для заданного входа, для функций потерь ранжирования высчитываются относительные расстояния между объектами, что еще называется обучением метрики. Согласно [8–10], данные функции потерь показывают сейчас наилучший результат в задачах обучения метрики.

В статье предлагается обучить энкодер (программный модуль, кодирующий картинку в вектор), основанный на глубокой нейронной сети, который смог бы различать объекты технического характера, часто используемые при задаче захвата объектов, тем самым предоставив возможность в дальнейшем построить детектор, в котором RPN-слой выдавал бы некоторое семантическое описание длины k для каждого предполагаемого кандидата в распознанные объекты и сравнивать их семантические описания с изображениями-эталонами, тем самым решая задачу классификации без ограничений в фиксированном количестве классов для распознавания.

Набор данных. В качестве данных был выбран набор данных MVTec ITODD [11] для 3D-детектирования объектов и оценки позиции с фокусом на технические объекты и приложения. Он содержит 28 объектов, распределенных среди 800 видеосцен и промаркированных 3D-позициями в качестве истинного значения. Сцены сканированы двумя промышленными 3D-сканерами и тремя черно-белыми камерами, позволяя реализовать методы, которые работают как с 3D-моделями, так и с 2D-изображениями. Объекты были выбраны таким образом, чтобы покрыть весь диапазон объектов с различной степенью отражения, симметрии, сложности, детализированности, компактности и размера. На рисунке 1 приведены примеры 28 объектов, представленных в таком наборе данных.

Материалы и методы. Подход, рассматриваемый в этой статье, является продолжением работы [12], идея которой заключается в обнаружении объектов для выполнения логистических задач. Данный подход расширяется для задачи захвата антропоморфной руки-манипулятора, первоначально решая задачу обнаружения объектов нескольких типов и формы. Для нивелирования непостоянного количества классов было решено ввести подход с одномоментным обучением путем создания метрического пространства для определенного пласта объектов на изображении. Идея получения век-



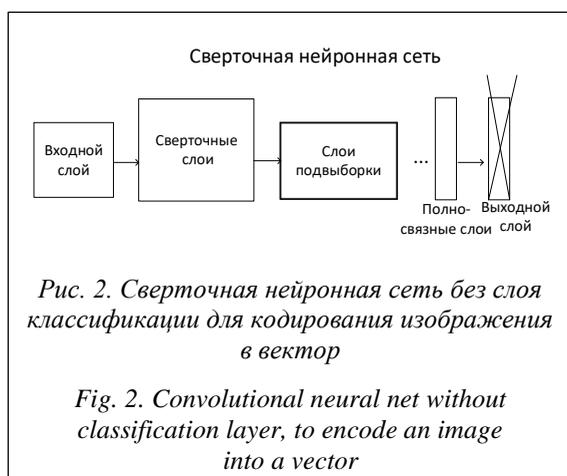
Рис. 1. Примеры объектов в датасете MVTec ITODD

Fig. 1. Examples of objects in dataset MVTec ITODD

тора, способного описать всю семантику изображения, очень проста. Для начала на выбранном наборе данных строится классификатор. Производится его обучение до достаточно хорошей точности, что позволяет получить модель, извлекающую из картинки некоторые семантические признаки, связанные с технической тематикой. В качестве классификатора выступает глубокая сверточная нейронная сеть с одной из уже проработанных архитектур. Это может быть архитектура семейств ResNet-50, ResNet-151, ResNet-81 или Inception v3 [13, 14]. Выбор архитектуры обусловлен размером сети, позволяющим вместиться в видеопамять, а дальнейший выбор делается исходя из перебора и подбора лучших параметров для получения лучшего качества.

Из полученного классификатора убирается слой классификации (рис. 2) и оставляется предпоследний слой, который является выходом нейронной сети и дает некоторое векторизованное представление изображения, содержащее семантическое описание объекта.

Чтобы данный слой имел свойство дистанции, то есть более близкие объекты с точки зрения косинусного, евклидова или манхэттенского расстояния были похожи по своему содержанию, а более дальние объекты не похожи, необходимо провести обучение по одной из метрик – Triplet Loss [8], Proxy-NCA [9] и Margin Loss [10], которые как раз и вводят меру дистанции между признаковыми описаниями изображений.



По полученному отображению из изображения в вектор строятся расстояния от изображений-кандидатов до изображений-эталонов, тем самым работа слоя классификатора выполняется вручную при помощи подсчета расстояния до изображений, с которыми мы хотим

сравнить потенциального кандидата и причислить к наиболее похожему. Для исключения ложноположительных ответов экспериментально подбирается минимальный порог для величины расстояния, чтобы причислить кандидата к одному из классов

Обучение. В качестве основы для извлечения признаков из картинки были взяты веса из модели, обученной на наборе данных Image Net Classification Challenge, где присутствуют объекты 1 000 различных классов, что позволяет нейронной сети еще до обучения извлекать некоторые общие признаки из изображений, поверх которых будет обучено отображение в метрическое пространство.

В качестве архитектур нейронной сети для эксперимента были взяты модели ResNet-50 и InceptionV3 как наиболее оптимальные по скорости и предсказательным возможностям. Для каждой из них был убран последний слой классификации и добавлен новый слой, размер которого является гиперпараметром D . Данный слой векторизации нейронной сети был инициализирован случайными маленькими числами из нормального распределения с нулевым средним и стандартным отклонением 0,01.

В качестве метода оптимизации выступил стохастический градиентный спуск. Длина градиентного шага – 0,001. После 10 000 итераций длина градиентного шага была снижена до 0,0001. Размер батчей (фрагментов набора данных, которые отправляются в видеокарту за одну итерацию) был равен 8 изображениям. Обучение проводилось на Amazon GPU сервере p2.xlarge с одной видеокарткой Nvidia Tesla K80 с 2 496 CUDA-ядрами и 12 GB видеопамью в течение 2 часов, скорость предсказания составляла около 100 мс.

Результаты. Для тестирования предложенной архитектуры были использованы модели ResNet-50 и InceptionV3 с тремя разными функциями потерь: Triplet loss, Proxy-NCA, Margin Loss. Для каждой комбинации были также рассмотрены разные размеры последнего слоя нейронной сети с точки зрения получения лучшего качества на тесте. Каждая комбинация оценивалась двумя метриками.

1. Recall@1 (Recall at One, или полнота среди первых членов выдачи) позволяет оценить алгоритмы, которые в качестве ответа возвращают последовательность различных объектов кандидатов. В случае с Recall@1 оценивается доля верно распознанных первых объектов в выдаче среди всех объектов данного класса.

2. Precision@1 (Precision at One, или точность среди первых членов выдачи) позволяет оценить алгоритмы, которые в качестве ответа на запрос возвращают последовательность различных объектов кандидатов. В случае с Precision@1 оценивается доля верно распознанных первых объектов в выдаче среди всех объектов, которые алгоритм отнес к данному классу.

Для данных метрик была построена таблица, отображающая результаты обучения различных комбинаций параметров, жирным шрифтом выделены оптимальные результаты (0,68 и 0,77).

Рассуждения. Полученные результаты позволяют удостовериться, что подход, схожий с применяемыми для распознавания лиц, хорошо работает и с объектами, которые используются в робототехнике в задачах захвата и оценки позиции объекта. Так как изображения черно-белые, предположительно, нейронной сетью были извлечены только признаки, связанные с формой объекта и типом поверхности, цвет в расчет не брался. Если учитывать цвет, то результаты могут быть улучшены, но черно-белый цвет был взят намеренно, так как в деталях, представленных в наборе данных, редко встречаются другие цвета. Многие объекты были расположены вперемешку, поэтому качество могло сильно ухудшаться, когда в обзор попадали несколько объектов. В дальнейшем данный подход будет также применен на повседневных предметах, где объекты одного класса могут сильно отличаться как формой и цветом, так и текстурой, что позволит значительно улучшить качество распознавания, а также добавить более точную постановку задачи для классификации, находя именно конкретный предмет.

Использование концепции мгновенного обучения в различных практических областях подробно описывается в обзоре [15].

Похожие на описываемые в данной статье наработки есть в исследовании [16]. Его авторы строят сквозную архитектуру, которая сразу сегментирует объекты для ранее не увиденных объектов. Но есть некоторые отличия:

- используется другой набор данных;
- используются изображения с глубиной (в данном исследовании – только черно-белые изображения);
- меньше размер вектора признаков, полученный в результате работы энкодера, что дает потенциальные возможности для более быстрой работы;
- сравнимое качество классификации при разных изначальных настройках;
- для обучения семантического описания объекта используются сиамские сети, в то время как в данном исследовании – сверточные нейронные сети с ранжированной функцией потерь без последнего слоя классификации.

Это говорит о заинтересованности в данной тематике других исследователей и актуальности набора технологий для реализации поставленной задачи.

В дальнейшем концепция, построенная в данной статье, способна реализовать способ поиска того объекта, который показывает человек в интерактивном режиме, так как реализован некоторый ассоциативный механизм, где изображения, схожие по содержанию, имеют свойство расстояния в векторизованном виде.

Подход к построению отображения в метрическое пространство планируется внедрить в архитектуру детектора, тем самым сделав концепцию моментного обучения применимой и

Результаты обучения векторизации технических изображений

Training results in vectorization of technical images

| Тип архитектуры нейронной сети | Размер выходного вектора | Функция потерь | | | | | |
|--------------------------------|--------------------------|-------------------------|------|-------------|-------------|-------------|------|
| | | Triplet loss | | Proxy-NCA | | Margin loss | |
| | | Метрика оценки качества | | | | | |
| | | R@1 | P@1 | R@1 | P@1 | R@1 | P@1 |
| ResNet 50 | D = 64 | 0,51 | 0,65 | 0,49 | 0,61 | 0,53 | 0,62 |
| | D = 128 | 0,65 | 0,72 | 0,68 | 0,77 | 0,59 | 0,65 |
| | D = 256 | 0,61 | 0,63 | 0,52 | 0,68 | 0,59 | 0,7 |
| | D = 512 | 0,45 | 0,7 | 0,39 | 0,81 | 0,42 | 0,67 |
| Inception V3 | D = 64 | 0,42 | 0,55 | 0,32 | 0,44 | 0,37 | 0,5 |
| | D = 128 | 0,49 | 0,59 | 0,55 | 0,63 | 0,54 | 0,61 |
| | D = 256 | 0,34 | 0,67 | 0,4 | 0,55 | 0,43 | 0,59 |
| | D = 512 | 0,65 | 0,7 | 0,49 | 0,68 | 0,51 | 0,67 |

для алгоритмов обнаружения. Далее это позволит строить алгоритмы, способные находить предварительно не заявленные объекты и оценивать их 3D- или 6D-положение. После этого можно решать задачу планирования расположения кисти манипулятора робота для захвата этого предмета в соответствии с методикой, описанной в [17].

Заключение. Был разработан алгоритм отображения изображений с техническими объектами в метрическое пространство, который позволяет проводить классификацию объектов по изображению-эталону, добавленному в систему после обучения сети.

Эксперимент с классификацией найденных объектов показал, что данный подход работает и дает результаты, схожие с обычной классификацией объектов. Это позволяет работать системе с нефиксированным количеством классов.

В дальнейшем планируется внедрить данный подход в архитектуру детектора в качестве отдельной ветви с собственными слоями, где будет оптимизироваться расстояние между семантическими представлениями найденных детектором изображений кандидатов, что позволит реализовать детектор с нефиксированным количеством классов.

Литература

1. Girshick R., Donahue J., Darrell T., Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2013. URL: <https://arxiv.org/abs/1311.2524> (дата обращения: 31.10.2019).
2. Girshick R. Fast R-CNN. 2015. URL: <https://arxiv.org/abs/1504.08083> (дата обращения: 31.10.2019)
3. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 2015. URL: <https://arxiv.org/abs/1506.01497> (дата обращения: 31.10.2019).
4. He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN. 2017. URL: <https://arxiv.org/abs/1703.06870> (дата обращения: 31.10.2019).
5. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. 2018. URL: <https://arxiv.org/abs/1804.02767> (дата обращения: 31.10.2019).
6. Lin T.-Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick C.L., Dollár P. Microsoft COCO: Common Objects in Context. Cornell University, 2014. URL: <https://arxiv.org/abs/1405.0312> (дата обращения: 31.10.2019).
7. Deng J., Dong W., Socher R., Li L., Li K., Li F. ImageNet: A large-scale hierarchical image database. Proc. IEEE Conf. CVPR, 2009. DOI: 10.1109/CVPR.2009.5206848.
8. Hermans A., Beyer L., Leibe B. In Defense of the Triplet Loss for Person Re-Identification. 2017. URL: <https://arxiv.org/abs/1703.07737> (дата обращения: 31.10.2019).
9. Movshovitz-Attias Y., Toshev A., Leung T.K., Ioffe S., Singh S. No Fuss Distance Metric Learning Using Proxies. 2017. URL: <https://arxiv.org/abs/1703.07464> (дата обращения: 31.10.2019).
10. Wei X., Wang H., Scotney B., Wan H. Minimum Margin Loss for Deep Face Recognition. 2018. URL: <https://arxiv.org/abs/1805.06741> (дата обращения: 31.10.2019). DOI: 10.1016/j.patcog.2019.107012.
11. Drost B., Ulrich M., Bergmann P., Härtinger P., Steger C. Introducing MVTEC ITODD – A dataset for 3D object recognition in industry. Proc. IEEE ICCVW, 2017. DOI: 10.1109/ICCVW.2017.257.
12. Ширкин А.Е., Тарачков М.В., Перминов И.К. Распознавание образов автономным роботом торгового зала // IV Всерос. конф. ГИСИС: сб. матер. Калининград, 2018. С. 382–387.
13. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. 2015. URL: <https://arxiv.org/abs/1512.03385> (дата обращения: 31.10.2019).
14. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions. 2014. URL: <https://arxiv.org/abs/1409.4842> (дата обращения: 31.10.2019).
15. Mahonya N., Campbella S., Carvalho A., Krpalkovaa L., Hernandez G.V., Harapanahallia S., Riordana D., Walsh J. One-shot learning for custom identification tasks. Procedia Manufacturing, 2019, vol. 38, pp. 186–193.
16. Schnieders B., Luo S., Palmer G., Tuyls K. Fully Convolutional One Shot Object Segmentation for Industrial Robotics. 2019. URL: <https://arxiv.org/abs/1903.00683> (дата обращения: 31.10.2019).
17. Тарачков М.В., Ширкин А.Е., Перминов И.К. Разработка элементов системы управления антропоморфным роботом AP-601 // Гибридные и синергетические интеллектуальные системы: матер. IV Всерос. конф. Калининград, 2018. С. 368–374.

Technical object projection in the image into metric space using deep neural networks for the detection problem

O.V. Tolstel¹, Ph.D. (Engineering), Associate Professor, tolstel.oleg@mail.ru

A.E. Shirkin¹, Postgraduate Student, shirkin39@mail.ru

A.L. Kalabin², Dr.Sc. (Physics and Mathematics), Professor, akalabin@yandex.ru

¹ Immanuel Kant Baltic Federal University, Institute of Physical and Mathematical Sciences and Information Technologies, Kaliningrad, 236041, Russian Federation

² Tver State Technical University, Tver, 170026, Russian Federation

Abstract. The paper presents the image vectorization algorithm containing technical objects. As technical objects are machine-building parts, fasteners, hardware. Vectorization refers to the transformation of an image into a vector for which the Euclidean distance has semantic meaning. This algorithm was created to improve the system of assessing the object position, where there is a problem of a variable number of types of objects for recognition.

The author proposes an approach to the formation of a metric space for images, where the image transformed into a vector by the metric l_2 can be compared with the image by the standard, thereby solving the problem of a non-constant number of classes. To add a new class, it is enough to add a reference image represented as a vector to the system and find the distance to it. If it is smaller than other images, then this reference will represent the type of object that was submitted to the input system. This approach is implemented in deep neural networks, where the last layer is removed and the penultimate layer is left, which represents the upper level of features extracted from the image. Such a neural network undergoes a learning process using the Triplet loss function, teaching the neural network to vectorize the image into metric space.

The program implementing the proposed algorithm is developed in Python 3.6 using the Jupyter Lab integrated environment for the Ubuntu 18.04 operating system.

The results of the experiment on the use of the proposed algorithm, which consisted in attributing the obtained images to a particular image - standard, are presented. To assess the quality of the algorithm, ranking metrics were used for search problems, where only the very first object in the list of nearest objects is evaluated.

The developed algorithm can be used for technical vision systems for robotic manipulators, and in the future this algorithm will be used as part of the control system for capturing objects by a robotic manipulator.

Keywords: deep neural networks, metric space, pattern recognition, object capture, robotic arms.

References

1. Girshick R., Donahue J., Darrell T., Malik J. *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*. 2013. Available at: <https://arxiv.org/abs/1311.2524> (accessed October 31, 2019).
2. Girshick R. *Fast R-CNN*. 2015. Available at: <https://arxiv.org/abs/1504.08083> (accessed October 31, 2019).
3. Ren S., He K., Girshick R., Sun J. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. 2015. Available at: <https://arxiv.org/abs/1506.01497> (accessed October 31, 2019).
4. He K., Gkioxari G., Dollár P., Girshick R. *Mask R-CNN*. 2017. Available at: <https://arxiv.org/abs/1703.06870> (accessed October 31, 2019).
5. Redmon J., Farhadi A. *YOLOv3: An Incremental Improvement*. 2018. Available at: <https://arxiv.org/abs/1804.02767> (accessed October 31, 2019).
6. Lin T.-Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick C.L., Dollár P. *Microsoft COCO: Common Objects in Context*. 2014. Available at: <https://arxiv.org/abs/1405.0312> (accessed October 31, 2019).
7. Deng J., Dong W., Socher R., Li L., Li K., Li F. ImageNet: A large-scale hierarchical image database. *Proc. IEEE Conf. CVPR*, 2009. DOI: 10.1109/CVPR.2009.5206848.
8. Hermans A., Beyer L., Leibe B. *In Defense of the Triplet Loss for Person Re-Identification*. 2017. Available at: <https://arxiv.org/abs/1703.07737> (accessed October 31, 2019).
9. Movshovitz-Attias Y., Toshev A., Leung T.K., Ioffe S., Singh S. *No Fuss Distance Metric Learning Using Proxies*. 2017. Available at: <https://arxiv.org/abs/1703.07464> (accessed October 31, 2019).
10. Wei X., Wang H., Scotney B., Wan H. *Minimum Margin Loss for Deep Face Recognition*. 2018. Available at: <https://arxiv.org/abs/1805.06741> (accessed October 31, 2019). DOI: 10.1016/j.patcog.2019.107012.

11. Drost B., Ulrich M., Bergmann P., Härtinger P., Steger C. Introducing MVTEC ITODD – A dataset for 3D object recognition in industry. *Proc. IEEE ICCVW*, 2017. DOI: 10.1109/ICCVW.2017.257.
12. Shirkin A.E., Tarachkov M.V., Perminov I.K. Image recognition by autonomous trading floor robot. *Proc. 4th All-Russ. Conf. GISIS-2018*, Kaliningrad, 2018, pp. 382–387 (in Russ.).
13. He K., Zhang X., Ren S., Sun J. *Deep Residual Learning for Image Recognition*. 2015. Available at: <https://arxiv.org/abs/1512.03385> (accessed October 31, 2019).
14. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. *Going Deeper with Convolutions*. 2014. Available at: <https://arxiv.org/abs/1409.4842> (accessed October 31, 2019).
15. Mahonya N., Campbella S., Carvalhoa A., Krpalkovaa L., Hernandez G.V., Harapanahallia S., Riordana D., Walsh J. One-shot learning for custom identification tasks. *Procedia Manufacturing*, 2019, vol. 38, pp. 186–193.
16. Schnieders B., Luo S., Palmer G., Tuyls K. *Fully Convolutional One Shot Object Segmentation for Industrial Robotics*. 2019. URL: <https://arxiv.org/abs/1903.00683> (accessed October 31, 2019).
17. Tarachkov M.V., Shirkin A.E., Perminov I.K. Development of elements of the control system anthropomorphic robot AR-601. *Proc. 4th All-Russ. Conf. GISIS-2018*, Kaliningrad, 2018, pp. 368–374 (in Russ.).

Для цитирования

Толстель О.В., Ширкин А.Е., Калабин А.Л. Проецирование технических объектов на изображении в метрическое пространство при помощи глубоких нейронных сетей для задачи детектирования // Программные продукты и системы. 2020. Т. 33. № 3. С. 464–470. DOI: 10.15827/0236-235X.131.464-470.

For citation

Tolstel O.V., Shirkin A.E., Kalabin A.L. Technical object projection in the image into metric space using deep neural networks for the detection problem. *Software & Systems*, 2020, vol. 33, no. 3, pp. 464–470 (in Russ.). DOI: 10.15827/0236-235X.131.464-470.