

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ГЛУБОКОГО МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧЕ ОБНАРУЖЕНИЯ КУПОЛА СЛЕПОЙ КИШКИ НА ВИДЕОДАННЫХ КОЛОНОСКОПИЧЕСКОГО ИССЛЕДОВАНИЯ

В. В. Хрящев

Ярославский государственный университет имени П. Г. Демидова, Ярославль, Россия
v.khryashchev@uniyar.ac.ru

Аннотация. *Актуальность и цели.* Рассматривается задача контроля качества колоноскопических исследований кишечника на основе анализа соответствующих видеоданных. Для ее решения предлагается обнаруживать на видеопотоке кадры, содержащие изображения купола слепой кишки. Данные фрагменты в видеопотоке являются ключевой точкой в процессе проведения колоноскопического исследования. *Материалы и методы.* Для решения поставленной задачи используются методы и алгоритмы глубокого машинного обучения и компьютерного зрения. Предлагается алгоритм обнаружения области купола слепой кишки на видеоизображениях с использованием сверточной нейронной сети архитектуры YOLO. *Результаты.* Получены метрики оценки качества работы предложенного алгоритма с использованием тестовой базы изображений, содержащей 1561 кадр области купола слепой кишки. Проанализированы результаты работы нейросетевого алгоритма в сравнении с популярным подходом на основе нейросетевой архитектуры SSD. Предложенный нейросетевой алгоритм показывает устойчивую работу и превосходит аналоги по стандартным метрикам. *Выводы.* Результаты будут являться базой для построения на основе разработанного алгоритма модуля анализа видеопотока в реальной медицинской системе для проведения колоноскопических исследований.

Ключевые слова: колоноскопические видеоизображения, купол слепой кишки, глубокое обучение, нейросетевой детектор объектов

Благодарности: автор работы выражает благодарность коллективу эндоскопического отделения ГБУЗ «Ярославская областная клиническая онкологическая больница» под руководством С. В. Кашина и Д. В. Завьялова за предоставленные видеофрагменты эндоскопических исследований, а также коллективу группы по разметке изображений под руководством А. С. Андержановой за помощь в проведении данного исследования.

Финансирование: работа выполнена в рамках реализации Программы развития ЯрГУ на период с 2021 по 2030 г.

Для цитирования: Хрящев В. В. Использование методов глубокого машинного обучения в задаче обнаружения купола слепой кишки на видеоданных колоноскопического исследования // Модели, системы, сети в экономике, технике, природе и обществе. 2023. № 4. С. 133–141. doi: 10.21685/2227-8486-2023-4-8

THE USE OF DEEP MACHINE LEARNING METHODS IN THE TASK OF DETECTING THE DOME OF THE CECUM ON THE VIDEO DATA OF COLONOSCOPIC EXAMINATION

V.V. Khryashchev

Yaroslavl State University named after P.G. Demidov, Yaroslavl, Russia
v.khryashchev@uniyar.ac.ru

Abstract. *Background.* Quality control of colonoscopic examinations of the intestine based on the analysis of relevant videodata is considered. To solve this problem, it is proposed to detect frames on the video stream containing images of the dome of the cecum. These fragments in the video stream are a key point in the process of conducting a colonoscopy examination. *Materials and methods.* To solve this problem, methods and algorithms of deep machine learning and computer vision are used. An algorithm is proposed for detecting the cecal dome region in video images using a convolutional neural network of the Yolo architecture. *Results.* Metrics for assessing the quality of work of the proposed algorithm were obtained using a test image database containing 1561 frames of the cecal dome region. The results of the neural network algorithm are analyzed in comparison with the popular approach based on the neural network architecture of SSD. The proposed neural network algorithm shows stable operation and outperforms its analogues in terms of standard metrics. *Conclusions.* The results will be the basis for building, based on the developed algorithm, a videostream analysis module in a real medical system for conducting colonoscopy studies.

Keywords: colonoscopic video images, cecum localization, deep learning, neural network object detection

Acknowledgements: the author of the work expresses gratitude to the team of the endoscopy department Yaroslavl Regional Clinical Oncology Hospital under the leadership of S.V. Kashin and D.V. Zavyalov for providing video clips of endoscopic studies, as well as to the team of the image marking group under the leadership of A.S. Anderzhanova for assistance in conducting this study.

Financing: the work was carried out within the framework of the implementation of the YarSU Development Program for the period 2021–2030.

For citation: Khryashchev V.V. The use of deep machine learning methods in the task of detecting the dome of the cecum on the video data of colonoscopic examination. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve* = *Models, systems, networks in economics, technology, nature and society*. 2023;(4):133–141. (In Russ.). doi: 10.21685/2227-8486-2023-4-8

Введение

На данном этапе развития технологии анализа видеоизображений одной из важнейших областей приложения разрабатываемых алгоритмов остается сфера медицины. Растущей областью подобных исследований является эндоскопия, в частности исследования желудка и кишечника с целью ранней диагностики онкологических заболеваний [1–3]. Процедура колоноскопии позволяет выявить доброкачественные и злокачественные образования в толстой кишке на ранней стадии их развития. О полноте такого исследования свидетельствует доведение эндоскопа до области слепой кишки пациента. Наличие такой области на записанном видеофайле свидетельствует о соблюдении проведенного исследования соответствующему регламенту [4–6].

Целью работы является автоматическое обнаружение области купола червеобразного отростка слепой кишки при проведении колоноскопического исследования. Решение такой задачи позволит построить важный элемент в системе контроля качества соответствующей медицинской процедуры [7, 8].

Материалы и методы

Для построения алгоритмов анализа видеопотока выбран подход на основе методов глубокого машинного обучения, показывающий свою эффективность в различных задачах анализа эндоскопических видеоизображений. В рамках исследования выбраны последние версии нейросетевых детекторов объектов архитектуры YOLO, наиболее популярные варианты YOLOv5, YOLOv7 и YOLOv8 [9]. Для сравнения алгоритмов в работе использовались также сверточные нейронные сети (СНС) популярной архитектуры SSD300 и SSD512 с дополнительным модулем VGG16 [10].

Рассмотрим общую архитектуру такой нейронной сети на примере YOLOv7. Процесс обработки эндоскопического изображения на основе алгоритма, использующего архитектуру YOLOv7, представлен на рис. 1.

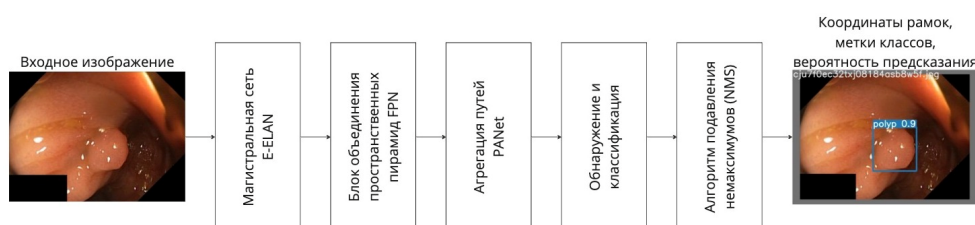


Рис. 1. Алгоритм анализа изображения на основе нейронной сети YOLOv7

Нейронная сеть семейства YOLO состоит из трех частей: магистральной части, части объединения пространственных пирамид и части предсказания.

В части сети, служащей для извлечения признаков объектов на входном изображении, используется СНС с использованием блоков E-ELAN (Extended Efficient Layer Aggregation Network). Соответствующий блок контролирует кратчайший и самый длинный градиентный путь для лучшей сходимости модели, используя операции расширения, перетасовки и слияния. Входные данные в нем подаются непосредственно через блок свертки 1×1 , в то время как два других подключения к блоку получены путем свертки входных данных двумя блоками свертки 3×3 с тем же канальным множителем. Далее все функции объединены и для получения полной информации также применяется свертка 1×1 .

К магистральной части добавляется часть объединения пространственных пирамид, выделяющая наиболее важные характерные особенности с разных уровней магистрали через сверточные слои блоков CBS (Convolution, BatchNormalization and SiLU) и SPPFCSPC (Spatial Pyramid Pooling Fast and Cross-Stage Partial Channel). Блок CBS является комбинацией сверточного слоя с разным размером ядра (1×1 , 3×1 и 3×2), слоя BatchNormalization (BN) и слоя активации с функциями типа SiLU (Sigmoid Linear Units). В конечных блоках пирамиды признаков свертки заменены блоками RepConv, необходимых для обновления параметров модели. Слои SPPFCSPC и RepConv являются

комбинацией слоев CBS с разным размером ядра для слияния признаков на разных масштабах, при этом слой SPPFCSPC собирает признаки и изображения, а слой RepConv производит их слияние в части формирования предсказания сети.

Далее следуют слои обнаружения и классификации, дающие итоговое предсказание модели. Тем самым на этапе обучения используются признаки в промежуточных слоях, улучшающие итоговую предсказательную способность модели. На выходе алгоритма традиционно выдаются координаты прямоугольной рамки, ограничивающей область интереса, метки соответствующих классов и вероятность предсказания, как показано на рис. 1.

Для эффективного обучения описанной современной архитектуры нейронной сети YOLO на практике требуются большие базы изображений (обычно от 10 000 размеченных объектов). К таким базам относятся широко известные наборы ImageNet, Pascal VOC, MS-COCO [11]. В области эндоскопии подобные проекты по сбору и разметке баз изображений все еще находятся в начальной стадии развития. Распространению открытых наборов таких данных мешают требования о сохранении персональной информации, а также сложность в разметке подобных видеоклипов, требующей высокой квалификации разметчика и валидации результатов с врачом экспертной квалификации. В связи с этим вопросы сбора, разметки и обмена подобными базами изображений медицинского профиля являются крайне важными для построения информационных систем в области цифровой медицины в целом и в эндоскопии в частности.

В данном исследовании сформирована база изображений, содержащих купол слепой кишки на основе размеченных видеоданных колоноскопических исследований, проведенных в эндоскопическом отделении ГБУЗ «Ярославская областная клиническая онкологическая больница». Использовались видеопотоки с эндоскопических систем OLYMPUS и PENTAX.

Записаны и обработаны видеоданные 21 исследования, длительность которых составляла 8–15 минут. Всего собрана 1561 пара изображений с разрешением 640×640 пикселей. Пару составляют исходное изображение без разметки и изображение с ее наличием (рис. 2). Традиционно база разделяется случайным образом на обучающий (тренировочный), тестовый и валидационный наборы. Валидационный набор используется для настройки, оптимизации и проверки точности процесса обучения СНС. Соответственно, пары изображений распределены между указанными наборами следующим образом: обучающий – 1228; тестовый – 159; валидационный – 174.

Как можно заметить из анализа изображений, приведенных на рис. 2, разметка области купола слепой кишки прямоугольной областью или попиксельная разметка (для обучения алгоритмов сегментации) представляет собой нетривиальный процесс и осуществляется обученным специалистом под контролем опытного врача-эндоскописта. В последнее время для сокращения временных затрат также используются полуавтоматические методы разметки медицинских изображений. В любом случае ни один из них не обходится без валидации итоговых данных медицинским персоналом экспертного уровня.

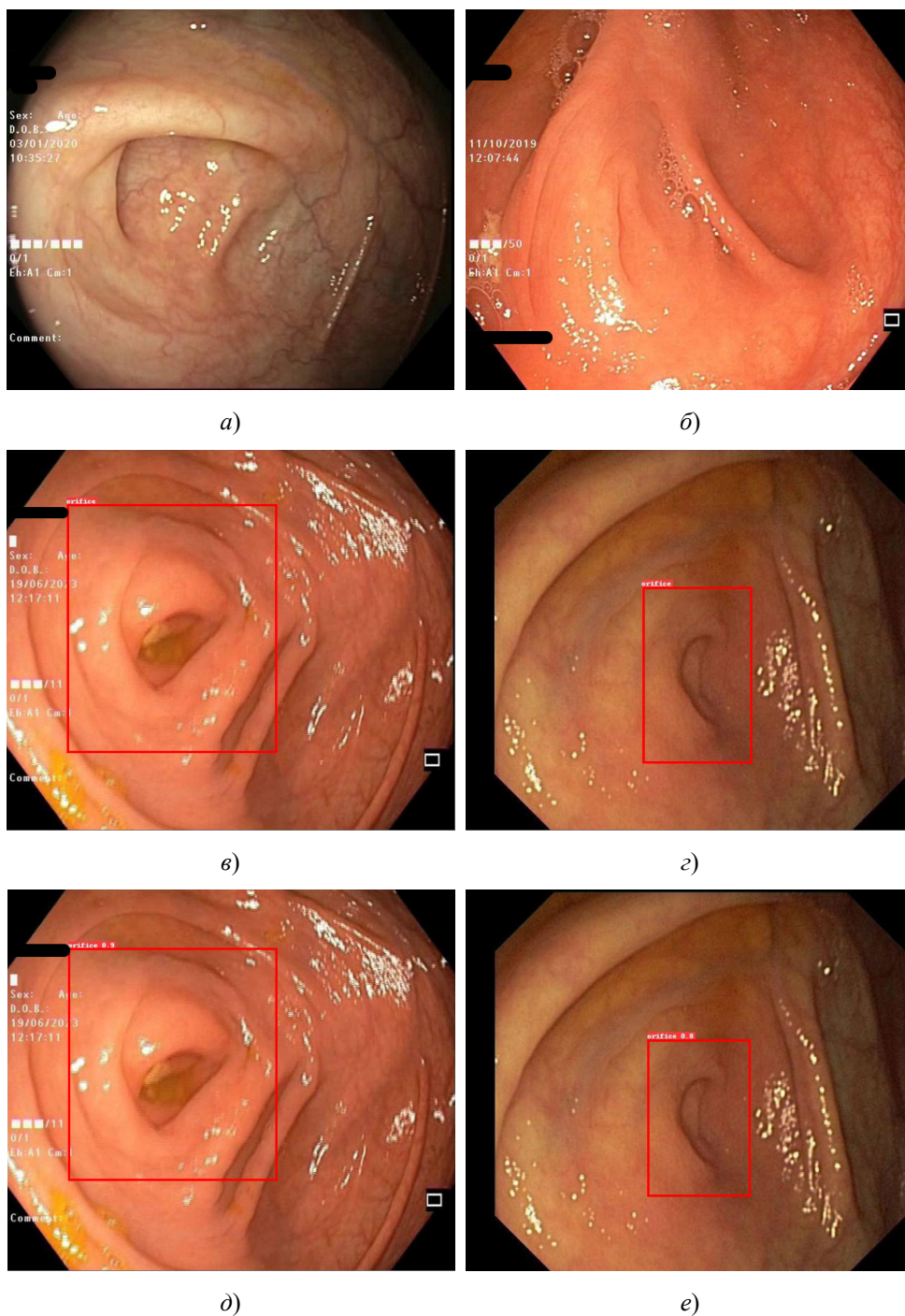


Рис. 2. Примеры изображений купола слепой кишки: а, б – без разметки; в, г – с разметкой экспертом; д, е – с разметкой нейросетевым алгоритмом

Результаты и обсуждение

Динамика ошибок TL (графики слева) и VL (графики справа) для соответствующих процедур обучения и валидации в зависимости от номера эпохи

приведена на рис. 3. Анализ этих зависимостей показывает, что процесс обучения СНС архитектур YOLO и SSD300 происходит устойчиво, сходясь в среднем на 150–200 итерациях. Для сети SSD512 заметны пульсации ошибок, происходящие на итерациях с номером до 75, далее наблюдается более плавный процесс их спада.

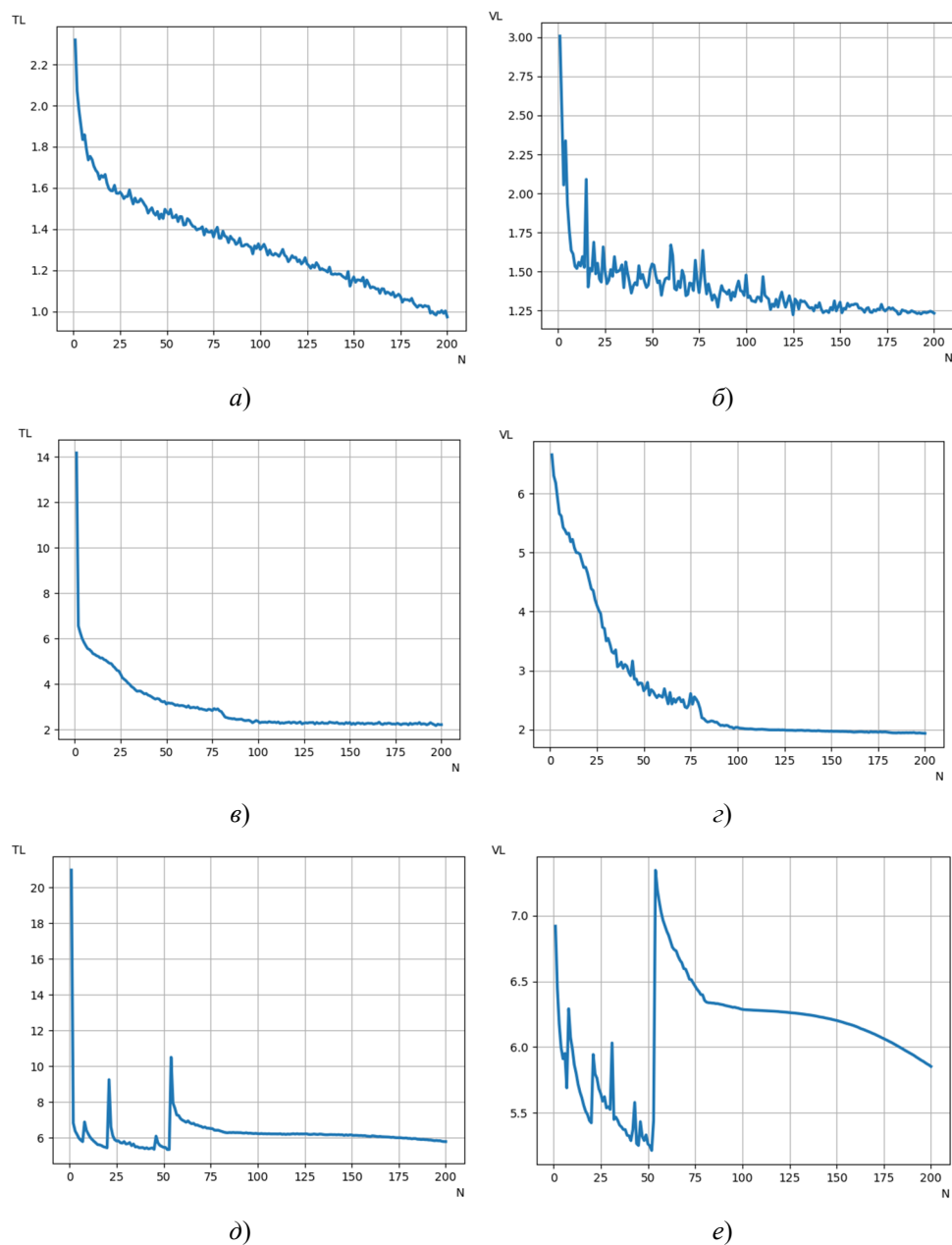


Рис. 3. Зависимость ошибки для процедур обучения и валидации от номера эпохи N:
 а, б – нейронная сеть YOLOv8s; в, г – нейронная сеть SSD300+VGG16; д, е –
 нейронная сеть SSD512+VGG16

Для обучения алгоритмов на основе архитектур YOLO использовалась последовательность (батч) из 16 изображений. В качестве оптимизационного алгоритма выбран алгоритм Adam (Adaptive Moment Estimation), сочетающий в себе как идею накопления движения, так и идею более слабого обновления весов для типичных признаков [12]. Начальная скорость обучения выбрана равной 10^{-3} , максимальное число эпох обучения до останова алгоритма – 250. Другие параметры алгоритма Adam составляли: $\beta_1 = 0,9$; $\beta_2 = 0,999$; $\epsilon = 10^{-8}$. На этапе обучения использовались функции потерь BCE с Logits Loss для сетей YOLOv5, YOLOv7 и YOLOR, а также BBox (IoU) для сети YOLOv8.

Для сравнения качества работы обученных нейросетевых алгоритмов обнаружения купола слепой кишки использовались стандартные метрики mAP50 и mAP50-95. Вторая из них рассчитывается как среднее из 10 значений mAP при 10 различных порогах IoU в интервале от 0,50 до 0,95 с шагом 0,05 [4].

В табл. 1 приведены результаты тестирования алгоритмов на всех исследуемых нейросетевых архитектурах. Все алгоритмы на базе архитектуры YOLO показывают практически одинаковый результат по метрике mAP50, при этом их преимущество по сравнению с аналогичным алгоритмом на основе архитектуры SSD весьма значительно и составляет 27–29 %.

Интересны также результаты по метрике mAP50-95. Здесь можно заметить отставание для алгоритмов на основе SSD в 28–30 %. Среди представленных алгоритмов детектирования на основе архитектуры YOLO лучшие результаты показал алгоритм YOLOv8s. Значение метрики mAP50 для него достигает 0,995, а усредненная по диапазону порогов метрика mAP50-95 – 0,698, что говорит о высокой способности детектирования объектов в широком диапазоне порогов детектирования. Другие модификации алгоритмов на основе архитектур YOLO 8-го поколения показывают близкие результаты. Практический выбор между данными архитектурами зависит от размера обучающей базы и требований к скорости работы итоговой обученной системы на фиксированной аппаратной базе. Преимущество этих архитектур по метрике mAP50-95 по сравнению с использованием архитектур YOLO пятого и седьмого поколения составляет 5–10 %.

Таблица 1

Тестирование нейросетевых алгоритмов детектирования

Архитектура сети/метрика	mAP50	mAP50-95
SSD300+VGG16	0,722	0,415
SSD512+VGG16	0,705	0,397
YOLOv5s	0,995	0,648
YOLOv7	0,997	0,607
YOLOv7-x	0,997	0,591
YOLOv8n	0,995	0,689
YOLOv8s	0,995	0,698
YOLOv8m	0,995	0,683
YOLOv8l	0,995	0,679

Таким образом, полученные результаты показывают преимущество использования нейросетевой архитектуры YOLOv8 для исследуемой задачи обнаружения купола слепой кишки на видеопотоке колоноскопического исследования.

Заключение

По результатам сравнения алгоритмов, приведенных в настоящем исследовании в области обнаружения купола слепой кишки на колоноскопических изображениях, можно сделать следующий вывод: рассмотренный алгоритм на базе архитектуры YOLOv8s превосходит приведенные в более ранних исследованиях аналоги по метрике mAP со всеми порогами. Исследования, проведенные в работе, послужат базой для построения на основе разработанного нейросетевого алгоритма обнаружения купола слепой кишки модуля анализа видеопотока в реальной эндоскопической системе.

Список литературы

1. Zhou S. K., Greenspan H., Shen D. *Deep Learning for Medical Image Analysis*. Elsevier Science, 2017.
2. Лебедев А. А., Хрящев В. В., Казина Е. М. [и др.]. Распознавание устья червеобразного отростка на эндоскопических изображениях прямой кишки на основе сверточной нейронной сети // Цифровая обработка сигналов и ее применение (DSPA-2020) : докл. 22-й Междунар. конф. М., 2020. С. 638–642.
3. Куваев Р. О., Никонов Е. Л., Кашин С. В. [и др.]. Контроль качества эндоскопических исследований, перспективы автоматизированного анализа эндоскопических изображений // Кремлевская медицина. Клинический вестник. 2013. Т. 2. С. 51–56.
4. Cho M., Kim J. H., Hong K. S. [et al.]. Identification of cecum time-location in a colonoscopy video by deep learning analysis of colonoscope movement // PeerJ. 2019. doi: 10.7717/peerj.7256
5. Хрящев В. В., Завьялов Д. В., Андержанова А. С. Классификация эндоскопических изображений устья червеобразного отростка на основе методов глубокого машинного обучения // Цифровая обработка сигналов. 2023. № 1. С. 35–38.
6. Akbar M., Shima R., Soroushmehr S. M. R. [et al.]. Classification of Informative Frames in Colonoscopy Videos Using Convolutional Neural Networks with Binarized Weights // Conference Paper in Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018. P. 65–68.
7. Bisschops R., Areia M., Coron E. [et al.]. Performance measures for upper gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative // Endoscopy. 2016. Vol. 48. P. 843–864.
8. Leiman D. A., Metz D. C., Ginsberg G. G. [et al.]. A novel electronic medical record – based workflow to measure and report colonoscopy quality measures // Clinical Gastroenterology and Hepatology. 2016. Vol. 14. P. 333337. doi: 10.1016/j.cgh.2015.12.001
9. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. М. : ДМК-Пресс, 2017. 652 с.
10. Liu W., Anguelov D., Erhan D. [et al.]. SSD: Single Shot MultiBox Detector // European conference on computer vision. 2016. P. 21–37.
11. Russakovsky O., Deng J., Su H. [et al.]. ImageNet Large Scale Visual Recognition Challenge // Computer Vision and Pattern Recognition. 2015.
12. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization // Machine Learning. 2017.

References

1. Zhou S.K., Greenspan H., Shen D. *Deep Learning for Medical Image Analysis*. Elsevier Science, 2017.

2. Lebedev A.A., Khryashchev V.V., Kazina E.M. et al. Recognition of the mouth of the vermiform process on endoscopic images of the rectum based on a convolutional neural network. *Tsifrovaya obrabotka signalov i ee primeneniye (DSPA-2020): dokl. 22-y Mezhdunar. konf. = Digital signal processing and its application (DSPA-2020) : reports of the 22nd International conference*. Moscow, 2020:638–642. (In Russ.)
3. Kuvaev R.O., Nikonov E.L., Kashin S.V. et al. Quality control of endoscopic studies, prospects for automated analysis of endoscopic images. *Kremlevskaya meditsina. Klinicheskiy vestnik = Kremlin medicine. Clinical Bulletin*. 2013;2:51–56. (In Russ.)
4. Cho M., Kim J.H., Hong K.S. et al. Identification of cecum time-location in a colonoscopy video by deep learning analysis of colonoscopy movement. *PeerJ*. 2019. doi: 10.7717/peerj.7256
5. Khryashchev V.V., Zav'yalov D.V., Anderzhanova A.S. Classification of endoscopic images of the mouth of the appendix on the basis of deep machine learning methods. *Tsifrovaya obrabotka signalov = Digital signal processing*. 2023;(1):35–38. (In Russ.)
6. Akbar M., Shima R., Soroushmehr S.M.R. et al. Classification of Informative Frames in Colonoscopy Videos Using Convolutional Neural Networks with Binarized Weights. *Conference Paper in Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2018:65–68.
7. Bisschops R., Areia M., Coron E. et al. Performance measures for upper gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. *Endoscopy*. 2016;48:843–864.
8. Leiman D.A., Metz D.C., Ginsberg G.G. et al. A novel electronic medical record – based workflow to measure and report colonoscopy quality measures. *Clinical Gastroenterology and Hepatology*. 2016;14:333337. doi: 10.1016/j.cgh.2015.12.001
9. Gudfellou Ya., Bendzhio I., Kurvill' A. *Glubokoe obucheniye = Deep learning*. Moscow: DMK Press, 2017:652. (In Russ.)
10. Liu W., Anguelov D., Erhan D. et al. SSD: Single Shot MultiBox Detector. *European conference on computer vision*. 2016:21–37.
11. Russakovsky O., Deng J., Su H. et al. ImageNet Large Scale Visual Recognition Challenge. *Computer Vision and Pattern Recognition*. 2015.
12. Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization. *Machine Learning*. 2017.

Информация об авторах / Information about the authors

Владимир Вячеславович Хрящев

кандидат технических наук, доцент,
доцент кафедры цифровых технологий
и машинного обучения,
Ярославский государственный
университет имени П. Г. Демидова
(Россия, г. Ярославль, ул. Советская, 14)
E-mail: v.khryashchev@uniyar.ac.ru

Vladimir V. Khryashchev

Candidate of technical sciences, associate
professor, associate professor
of the sub-department of digital
technologies and machine learning,
Yaroslavl State University
named after P.G. Demidov
(14 Sovetskaya street, Yaroslavl, Russia)

Автор заявляет об отсутствии конфликта интересов /

The author declares no conflicts of interests.

Поступила в редакцию/Received 15.05.2023

Поступила после рецензирования/Revised 15.09.2023

Принята к публикации/Accepted 05.10.2023