Оценка эффективности современных архитектур глубокого обучения в задаче сегментации трещин

Архипов Вадим Олегович
магистрант, Сибирский государственный университет геосистем и технологий,

Шарапов Артем Андреевич
ст. преподаватель кафедры прикладной информатики и информационных систем, Сибирский государственный университет геосистем и технологий, arhipov-vo2024@sgugit.ru

В работе исследуется применение алгоритмов глубокого обучения для сегментации трещин на фасадах зданий и сооружений. Актуальность исследования обусловлена необходимостью повышения объективности мониторинга технического состояния конструкций и снижения субъективности при визуальных инспекциях. Цель работы – разработка и экспериментальная апробация технологического подхода к автоматизированному выявлению и количественной оценке дефектов по фотографическим данным. В ходе исследования сформирован размеченный набор данных из 7173 изображений, обучены и сравнены архитектуры YOLO (версии 11 и 26), Mask R-CNN и Mask2Former с бэкбонами ResNet-50 и Swin-T. Наилучшие формальные показатели достигнуты моделью YOLO11x (mAP50 = 0,7177 для детекции, 0,6927 для сегментации), однако при визуальной оценке на реальных данных наиболее высокую полноту обнаружения дефектов показала модель Mask2Former Swin-T. Теоретическая значимость состоит в сравнительном анализе нейросетевых архитектур применительно к задаче сегментации малоразмерных объектов на сложном фоне. Практическая ценность заключается в возможности интеграции обученных моделей в системы цифрового мониторинга зданий и геоинформационные системы для автоматизированной фиксации дефектов и поддержки принятия управленческих решений при технической эксплуатации объектов.

Ключевые слова: техническое состояние зданий и сооружений, обнаружение деформаций, компьютерное зрение, глубокое обучение, сегментация изображений.

Введение

Анализ аварийных ситуаций, возникающих в процессе эксплуатации зданий и сооружений, показывает, что человеческий фактор часто приводит к игнорированию или сокрытию критических деформаций. Зачастую деформации фиксируются проверяющими лицами лишь формально, а их реальная динамика не отслеживается. Также, существующие методы визуального контроля не обеспечивают количественной оценки параметров дефектов и характеризуются высокой субъективностью [1–3].

Современные методы компьютерного зрения и глубокого обучения позволяют автоматизировать процессы выявления и анализа дефектов по цифровым изображениям. Однако существующие решения ограничены сложностями сегментации объектов с низкой плотностью пикселей, сложной геометрией и вариативным визуальным фоном, характерным для трещин строительных конструкций [4].

Целью работы является разработка и экспериментальная апробация метода автоматизированной сегментации трещин строительных конструкций на основе алгоритмов глубокого обучения.

Объектом исследования являются процессы мониторинга технического состояния строительных конструкций на основе анализа цифровых изображений.

Предметом исследования выступают методы и алгоритмы глубокого обучения для сегментации трещин на изображениях фасадов и несущих элементов зданий и сооружений.

Научная новизна исследования заключается в разработке подхода к обучению моделей сегментации для работы в условиях сложного визуального фона. Предложенный подход позволяет повысить точность выделения дефектов в условиях сложного визуального фона.

Теоретическая значимость работы состоит в анализе современных архитектур нейронных сетей на специфической задаче сегментации объектов с низкой плотностью пикселей и сложной геометрией. Это расширяет представления о применимости обученных моделей в техническом обследовании зданий и сооружений.

Практическая значимость заключается в разработке решения для автоматизированного выявления и дальнейшей оценки параметров деформаций зданий и сооружений. Результаты исследования могут быть использованы при создании систем цифрового мониторинга зданий, а также интегрированы в геоинформационные системы для пространственного анализа и учета дефектов строительных конструкций.

Сбор и предварительная обработка набора данных

Первым этапом исследования являлось формирование набора данных, предназначенного для обучения моделей сегментации трещин зданий и сооружений. Процесс подготовки включал поиск исходных изображений, их преобразование к единому виду и разметку согласно требованиям к обучению моделей. Исходные изображения фасадов и конструктивных элементов зданий отбирались из открытых источников в сети Интернет, а также из доступных к свободному использованию репозиториев.

Поскольку было необходимо выделять каждую трещину отдельно, чтобы на основе имеющихся данных можно было бы вычислить ее геометрию, использовались актуальные архитектуры, поддерживающие инстанс-сегментацию: YOLO, Mask RCNN и Mask2Former. (архитектуры на базе CNN и трансформеров).

Для обеспечения совместимости данных с рассматриваемыми архитектурами сверточных нейронных сетей была проведена техническая адаптация исходных материалов. Изображения высокого разрешения подвергались фрагментации и приведению к единому входному разрешению 800 на 800 пикселей. Чтобы избежать геометрических искажений и деформации структуры трещин при изменении разрешения, применялся метод дополнения кадра черными полями (padding). Это позволило сохранить исходное соотношение сторон объектов, что очень важно для корректного обучения алгоритмов сегментации [5–8].

Заключительная стадия подготовки данных заключалась в аннотировании изображений с использованием соответствующих инструментов. Разметка наборов данных, полученных из открытых источников была отредактирована и пересмотрена вручную.

Процесс ручной корректировки разметки изображений позволил устранить ошибки автоматической сегментации и уточнить границы мелких дефектов, что обеспечило высокую достоверность эталонных данных на последующих этапах обучения (рис. 1).



a) b) c)

Рис. 1. Примеры размеченных изображений

Итоговый размер набора данных составил 7173 изображения, из которых 5736 использовались для обучения моделей, а 1437 – для валидации. Разделение выборки осуществлялось с учетом обеспечения репрезентативности

данных по типам дефектов, условиям съемки и характеристикам объектов. Это позволило минимизировать смещение распределения признаков и обеспечить корректную оценку обобщающей способности обучаемых моделей.

Обучение моделей и анализ результатов

Для проведения экспериментов по обучению моделей сегментации использовались фреймворки Ultralytics и обновленная версия набора инструментов для обучения и использования моделей обнаружения и сегментации объектов mmdetection. Данный выбор обусловлен наличием в них набора моделей с предобученными весами, полученными на датасете COCO [9–11].

Управление версиями набора данных осуществлялось с применением платформы Roboflow Universe. Экспорт аннотаций выполнялся посредством Python SDK. Для обеспечения совместимости с выбранными фреймворками данные были приведены к форматам COCO JSON и YOLO.

Процессы конфигурирования и инициализации обучения в выбранных фреймворках имеют принципиальные различия в уровне абстракции и степени контроля над гиперпараметрами.

Фраймворк Ultralytics реализует высокоуровневый подход, основанный на инкапсуляции сложных настроек и минимизируя участие пользователя. Инициализация обучения осуществляется с использованием интерфейса командной строки (CLI) или высокоуровневого Python API. При этом автоматизируются процессы построения вычислительного графа модели и предустановки базовых стратегий аугментации данных, что ускоряет цикл разработки, но ограничивает возможности низкоуровневой модификации алгоритмов [12].

В отличии от этого MMDetection строится на принципах модульного подхода и требует детальной декомпозиции каждой модели через систему конфигурационных файлов. Для каждой архитектуры формируется индивидуальная конфигурация, описывающая не только структуру сети, но и специфику обработки данных на каждом этапе. Несмотря на повышенную сложность настройки, такой подход обеспечивает более полный контроль, позволяя изменять параметры компонентов и управлять процессом обучения [13].

Несмотря на различия в реализации, общий алгоритм подготовки к обучению в обоих фреймворках включает следующие этапы:

выбор архитектуры модели из набора предобученных решений;

задание путей к изображениям и файлам аннотаций в форматах YOLO или COCO;

настройка гиперпараметров обучения, а именно установка лимитов (количество эпох/итераций), размера пакета данных, а также выбор оптимизатора и стратегии изменения темпа обучения;

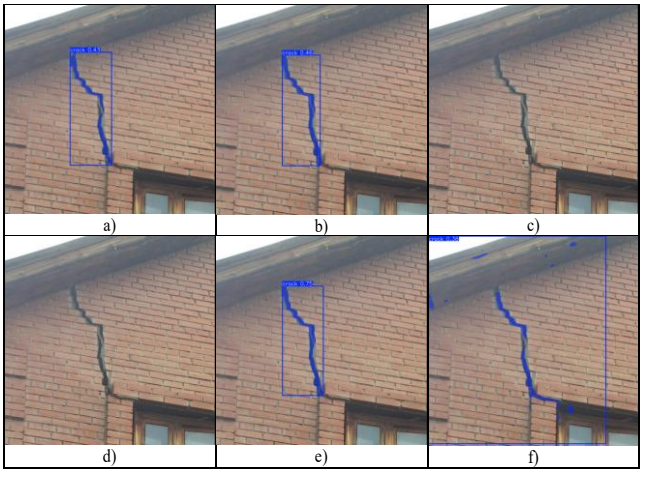
запуск процесса обучения на вычислительных ресурсах с последующим мониторингом метрик качества в режиме реального времени.

Обучение моделей выполнялось на графическом ускорителе NVIDIA GeForce RTX 4090 в среде JupyterHub с использованием официальных Docker-образов фреймворков. Результаты обучения представлены в табл. 1.

Таблица 1

Точность обученных моделей (mAP)

Модель	Bbox mAP50-95	Bbox mAP50	Segm mAP50-95	Segm mAP50
Mask rcnn r50	0,4470	0,6400	0,1320	0,3820
Mask2Former r50	0,3480	0,4900	0,1920	0,4680
Mask2Former swin-t	0,3650	0,5130	0,2050	0,4940
Yolo11x	0,5709	0,7177	0,3237	0,6927
Yolo11l	0,5540	0,7120	0,3037	0,6776
Yolo11m	0,5405	0,7139	0,3035	0,6729
Yolo26x	0,4879	0,6727	0,2515	0,6174
Yolo26l	0,4946	0,6835	0,2585	0,6216
Yolo26m	0,4978	0,6824	0,2606	0,6308



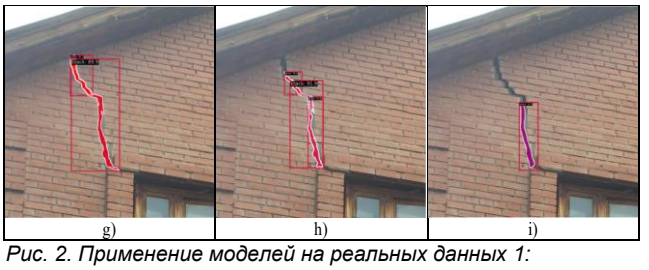


Рис. 2. Применение моделей на реальных данных 1:

Наивысшие формальные показатели точности продемонстрировала архитектура YOLO11x, достигнув значений mAP50 в 0,7177 для детекции и 0,6927 для сегментации. YOLO26 показала более низкие результаты по сравнению с YOLO11.

Модель Mask R-CNN r50 показала минимальную точность формирования масок среди всех рассматриваемых решений (Segm mAP50 = 0,3820).

Модели Mask2Former (с бэкбонами ResNet-50 и Swin-T) заняли промежуточное положение по метрикам mAP. При этом переход от сверточного бэкбона к трансформерному (Swin-T) позволил увеличить точность сегментации с 0,4680 до 0,4940 (mAP50). Кроме того, модель с бэкбоном Swin-T продемонстрировала наиболее высокую точность и полноту наложения масок при визуальном осмотре результатов сегментации на реальных данных (рис. 2, 3).

a) YOLO 26x; b) YOLO 26l; c) YOLO 26m; d) YOLO 11x; e) YOLO 11l; f) YOLO 11m; g) Mask2Former Swin-T; h) Mask2Former R50; i) Mask R-CNN R50

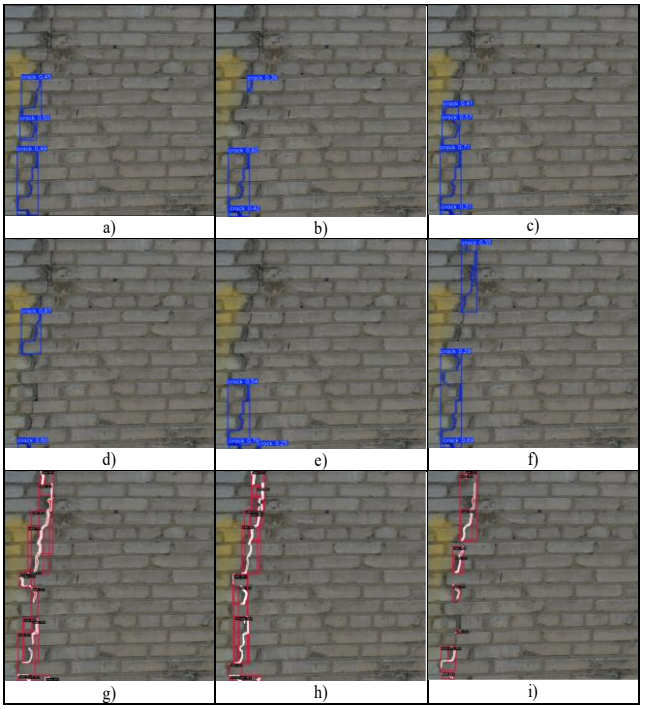


Рис. 3. Применение моделей на реальных данных 2:

a) YOLO 26x; b) YOLO 26l; c) YOLO 26m; d) YOLO 11x; e) YOLO 11l; f) YOLO 11m; g) Mask2Former Swin-T; h) Mask2Former R50; i) Mask R-CNN R50

Из проведённых экспериментов можно сделать вывод, что модель Mask2Former с архитектурой Swin-T демонстрирует более высокую полноту обнаружения дефектов на реальных данных по сравнению с другими рассматриваемыми моделями. Низкие значения mAP обусловлены фрагментацией протяженных сегментов трещин, что приводит к штрафам при оценке по метрике COCO. Однако данный эффект не снижает практическую применимость модели, поскольку выявление дефектов сохраняет достаточную точность для задач мониторинга.

Заключение

Результатом проведенного исследования является реализация полного цикла подготовки данных и обучения моделей глубокого обучения для сегментации дефектов строительных конструкций.

В результате был сформирован и верифицирован набор данных для обучения алгоритмов сегментации. В рамках данного этапа выполнена предварительная обработка изображений. Реализованы два независимых процесса обучения, учитывающих архитектурные особенности выбранных фреймворков.

В среде Ultralytics обучены модели семейства YOLOv26 и YOLOv11 для задач сегментации, которые в свою очередь характеризуются высокой скорость инференца и возможностью эффективной последующей оптимизации в форматы ONNX и TensorRT.

В среде MMDetection настроены и обучены более сложные архитектуры, включая двухэтапный детектор Mask R-CNN и трансформерную модель Mask2Former с использованием базовых архитектур Swin-T и ResNet-50 [14–16].

Проведённое исследование подтверждает эффективность применения современных архитектур глубокого обучения для автоматизации задач диагностики строительных конструкций. В результате получено комплексное решение, включающее подготовленный набор данных, настроенную вычислительную инфраструктуру и совокупность обученных моделей, пригодных для дальнейшего тестирования и интеграции в системы автоматизированного мониторинга технического состояния зданий, в том числе в составе геоинформационных систем.

Литература

1. В Оренбурге обрушилась школа // Эксперт-Урал : сетевое издание. – 2008. – 6 окт. – URL: https://expert-ural.com/archive/40-348/v-orenburge-obrushilas-shkola.html (дата обращения: 19.11.2025).
2. В обрушении дома по ул. Куйбышева, 103 оказались виноваты осадка фасада и кондиционер // Новости Перми : информационный портал. – 2015. – 25 авг. – URL: https://permnews.ru/novosti/society/2015/08/25/v_obrushenii_doma_po_ul__kuybysheva_103_okazalis_vinovaty_osadka_fasada_i_kondicioner/ (дата обращения: 19.11.2025).
3. В Петербурге эвакуировали дом на Большой Зелениной из-за угрозы обрушения // Собака.ru : журнал о людях в Петербурге. – 2024. – 4 авг. – URL: https://www.sobaka.ru/city/city/203634 (дата обращения: 19.11.2025).
4. Филонова Марьяна Андреевна, Широбокова Светлана Николаевна Сравнительный анализ моделей глубоких свёрточных сетей для сегментации дефектов в строительных конструкциях // Известия вузов. Северо-Кавказский регион. Серия: Технические науки. 2025. №. URL: https://cyberleninka.ru/article/n/sravnitelnyy-analiz-modeley-glubokih-svyortochnyh-setey-dlya-segmentatsii-defektov-v-stroitelnyh-konstruktsiyah (дата обращения: 19.11.2025).
5. Unel, F. O. The Power of Tiling for Small Object Detection / F. O. Unel, B. O. Ozkalayci, C. Cigla // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. – 2019. – P. 9–15.
6. Siriborvornratanakul, T. Downstream Semantic Segmentation Model for Low-Level Surface Crack Detection / T. Siriborvornratanakul // Advances in Multimedia. – 2022. – Vol. 2022. – Art. 3712289. – 12 p. – URL: https://doi.org/10.1155/2022/3712289 (дата обращения: 19.11.2025).
7. K. Ma, M. Hao, W. Shang // Study on the Influence of Label Image Accuracy on the Performance of Concrete Crack Segmentation Network Models // Sensors. – 2024. – Vol. 24, I. 4. – Art. 1068. – URL: https://doi.org/10.3390/s24041068 (дата обращения: 19.11.2025).
8. Шарден, Б. Крупномасштабное машинное обучение вместе с Python : учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти ; перевод с английского А. В. Логунова. – Москва : ДМК Пресс, 2018. – ISBN 978-5-97060-506-6. Текст : электронный // Лань : электронно-библиотечная система. URL: https://e.lanbook.com/book/105836 (дата обращения: 19.11.2025). Режим доступа: для авториз. пользователей.
9. Васильев М. Е., Шалимов А. С., Савина О. А. Обзор версий YOLO: одноэтапная модель сверточной нейронной сети // Universum: технические науки. 2025. №6 (135). URL: https://cyberleninka.ru/article/n/obzor-versiy-yolo-odnoetapnaya-model-svertochnoy-neyronnoy-seti (дата обращения: 19.11.2025).
10. Архипов, В. О. Обучение нейросети для поиска людей по фотоснимкам, полученным с БПЛА / В. О. Архипов, А. А. Шарапов // Интерэкспо Гео-Сибирь. – 2024. – Т. 7, № 1. – С. 29-34. – EDN PXEMPJ. URL: https://www.elibrary.ru/download/elibrary_69212979_44018838.pdf (Дата обращения 19.11.2025).
11. Архипов В. О. Автоматизация распознавания объектов на космических снимках методами глубокого обучения / В. О. Архипов. – Текст : электронный // Интерэкспо ГЕО-Сибирь : XXI Международный научный конгресс, 21–22 мая 2025 г., Новосибирск : сборник материалов в 8 т. Т. 6 : Магистерская научная сессия «Первые шаги в науке». – Новосибирск : СГУГиТ, 2025. – Т. 6, № 1. – с. 9-15. – DOI 10.33764/2618-981X-2025-6-1. URL: https://geosib.sgugit.ru/upload/geosibir/sborniki/2025/tom-6/009-015.pdf (Дата обращения 10.02.2026).
12. Ultralytics YOLO: Open-source computer vision models / Ultralytics. – 2023. – URL: https://github.com/ultralytics/ultralytics (дата обращения: 19.11.2025).
13. OneDL MMDetection: Open source object detection toolbox based on MMDetection / VBTI Development. – 2023. – URL: https://github.com/VBTI-development/onedl-mmdetection (дата обращения: 19.11.2025).
14. Cheng, B. Masked-attention Mask Transformer for Universal Image Segmentation / B. Cheng, I. Misra, A. G. Schwing [et al.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2022. – P. 1290–1299. – URL: https://arxiv.org/pdf/2112.01527 (дата обращения: 19.03.2026).
15. Redmon, J. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon, S. Divvala, R. Girshick [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 779–788. – URL: https://arxiv.org/pdf/1506.02640 (дата обращения: 19.03.2026).
16. He, K. Mask R-CNN / K. He, G. Gkioxari, P. Dollár [et al.] // Proceedings of the IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2961–2969. – URL: https://arxiv.org/pdf/1703.06870 (дата обращения: 19.03.2026).

Definition of the Characteristics of the Unmanned Aviation System when Carrying out Search and Rescue Operations in Wetted Areas

Arkhipov V.O., Sharapov A.A.
Siberian State University of Geosystems and Technologies

JEL classification: L61, L74, R53

This paper investigates the application of deep learning algorithms for crack segmentation on building facades and structural elements. The relevance of the study stems from the need to improve the objectivity of structural health monitoring and to reduce subjectivity inherent in visual inspections. The aim of the work is to develop and experimentally test a technological approach for automated defect detection and quantitative assessment based on photographic data. In the course of the research, a labeled dataset of 7,173 images was compiled, and several architectures were trained and compared, including YOLO (versions 11 and 26), Mask R-CNN, and Mask2Former with ResNet-50 and Swin-T backbones. The highest formal performance was achieved by YOLO11x (mAP50 = 0.7177 for detection and 0.6927 for segmentation); however, in visual evaluation on real-world data, the Mask2Former Swin-T model demonstrated the highest recall in defect detection. The theoretical significance lies in a comparative analysis of neural network architectures for segmenting small-scale objects against complex backgrounds. The practical value consists in the potential integration of the trained models into digital building monitoring systems and geographic information systems for automated defect recording and decision support in building maintenance and operation.

Keywords: structural health monitoring, deformation detection, computer vision, deep learning, image segmentation.

References

1. V Orenburge obrushilas' shkola [A school collapsed in Orenburg]. Ekspert-Ural: setevoe izdanie [Expert-Ural: online edition]. 2008, October 6. Available at: https://expert-ural.com/archive/40-348/v-orenburge-obrushilas-shkola.html (accessed November 19, 2025) [in Russian].
2. V obrushenii doma po ul. Kuybysheva, 103 okazalis' vinovaty osadka fasada i konditsioner [The collapse of the house on Kuybyshev St., 103 was caused by facade settlement and an air conditioner]. Novosti Permi: informatsionnyy portal [Perm News: information portal]. 2015, August 25. Available at: https://permnews.ru/novosti/society/2015/08/25/v_obrushenii_doma_po_ul__kuybysheva_103_okazalis_vinovaty_osadka_fasada_i_kondicioner/ (accessed November 19, 2025) [in Russian].
3. V Peterburge evakuirovali dom na Bol'shoy Zeleninoy iz-za ugrozy obrusheniya [A house on Bolshaya Zelenina was evacuated in St. Petersburg due to the threat of collapse]. Sobaka.ru: zhurnal o lyudyakh v Peterburge [Sobaka.ru: a magazine about people in St. Petersburg]. 2024, August 4. Available at: https://www.sobaka.ru/city/city/203634 (accessed November 19, 2025) [in Russian].
4. Filonova, M. A., & Shirobokova, S. N. (2025). Sravnitel'nyy analiz modeley glubokikh svyortochnykh setey dlya segmentatsii defektov v stroitel'nykh konstruktsiyakh [Comparative analysis of deep convolutional network models for defect segmentation in building structures]. Izvestiya vuzov. Severo-Kavkazskiy region. Seriya: Tekhnicheskie nauki [University News. North Caucasus Region. Technical Sciences Series], (6). Available at: https://cyberleninka.ru/article/n/sravnitelnyy-analiz-modeley-glubokih-svyortochnyh-setey-dlya-segmentatsii-defektov-v-stroitelnyh-konstruktsiyah (accessed November 19, 2025) [in Russian].
5. Unel, F. O., Ozkalayci, B. O., & Cigla, C. (2019). The Power of Tiling for Small Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 9–15.
6. Siriborvornratanakul, T. (2022). Downstream Semantic Segmentation Model for Low-Level Surface Crack Detection. Advances in Multimedia, 2022, Art. 3712289, 12 p. DOI 10.1155/2022/3712289 (accessed November 19, 2025).
7. Ma, K., Hao, M., & Shang, W. (2024). Study on the Influence of Label Image Accuracy on the Performance of Concrete Crack Segmentation Network Models. Sensors, 24(4), Art. 1068. DOI 10.3390/s24041068 (accessed November 19, 2025).
8. Sharden, B., Massaron, L., & Boskетti, A. (2018). Krupnomasshtabnoe mashinnoe obuchenie vmeste s Python [Large Scale Machine Learning with Python]. Moscow: DMK Press Publ., 978-5-97060-506-6. Available at: https://e.lanbook.com/book/105836 (accessed November 19, 2025) [in Russian].
9. Vasil'ev, M. E., Shalimov, A. S., & Savina, O. A. (2025). Obzor versiy YOLO: odnoetapnaya model' svertochnoy neyronnoy seti [YOLO versions overview: a single-stage convolutional neural network model]. Universum: tekhnicheskie nauki [Universum: Technical Sciences], 6(135). Available at: https://cyberleninka.ru/article/n/obzor-versiy-yolo-odnoetapnaya-model-svertochnoy-neyronnoy-seti (accessed November 19, 2025) [in Russian].
10. Arkhipov, V. O., & Sharapov, A. A. (2024). Obuchenie neyroseti dlya poiska lyudey po fotosnimkam, poluchennym s BPLA [Neural network training for searching people using UAV imagery]. Interekspo Geo-Sibir' [Interexpo GEO-Siberia], 7(1), 29–34. Available at: https://www.elibrary.ru/download/elibrary_69212979_44018838.pdf (accessed November 19, 2025) [in Russian].