
КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ COMPUTER SIMULATION HISTORY

УДК 004.932

DOI 10.52575/2687-0932-2023-50-3-624-632

Автоматизация формирования выборки изображений естественных сцен для обучения и тестирования нейронных сетей

Гусев А.В.

ООО «Техинтегратор»,
Россия, 117303, г. Москва, ул. Каховка, д. 11, стр. 1
E-mail: 9671999461@mail.ru

Аннотация. Для решения задач распознавания образов на изображениях активно применяются искусственные нейронные сети. Для их обучения необходимо подготавливать наборы размеченных данных для формирования обучающих и валидационных выборок. Формирование таких выборок «вручную» требует высокой квалификации специалистов и больших временных, а, следовательно, и материальных затрат. Данная работа посвящена формированию набора основных методов и технологий, а также созданию и апробированию архитектуры решения для автоматизированного сегментирования и аннотирования изображений естественных сцен с целью создания выборок для обучения и тестирования нейронных сетей.

Ключевые слова: нейронные сети, датасет, сегментирование изображений, аннотирование изображений, обучающая выборка

Для цитирования: Гусев А.В. 2023. Автоматизация формирования выборок изображений естественных сцен для обучения и тестирования нейронных сетей. Экономика. Информатика. 50(3): 624–632 . DOI: 10.52575/2687-0932-2023-50-3-624-632

Automation of Sampling of Images of Natural Scenes for Training and Testing Neural Networks

Aleksandr V. Gusev

Limited Liability Company «Techintegrator»,
11 Kahovka St, bldg 1, Moscow, 117303, Russian Federation
E-mail: 9671999461@mail.ru

Abstract. Artificial neural networks are actively used to solve problems of pattern recognition in images. For their training, it is necessary to prepare labeled data sets for the formation of training and validation samples. The formation of such samples "manually" requires highly qualified specialists and large time and, consequently, material costs. This work is devoted to the formation of a set of basic methods and technologies, as well as the creation and testing of the solution architecture for automated segmentation and annotation of images of natural scenes in order to create samples for training and testing neural networks.

Keywords: neural networks, dataset, image segmentation, image annotation, training set

For citation: Gusev A.V. 2023. Automation of Sampling of Images of Natural Scenes for Training and Testing Neural Networks. Economics. Information technologies, 50(3): 624–632 (in Russian). DOI: 10.52575/2687-0932-2023-50-3-624-632

Введение

Целью работы является формирование набора основных методов и технологий, а также создание и апробирование архитектуры решения для автоматизированного сегментирования и аннотирования изображений естественных сцен с целью создания выборок для обучения и тестирования нейронных сетей.

Основой для разработки автоматизированных решений по формированию указанных выборок является необходимость решать следующие проблемы:

- ручная разметка изображений (выделение элементов на изображениях, присвоение аннотации каждому элементу) является трудоемкой и высокозатратной по времени процедурой. Она может занимать порядка 80% временного ресурса в рамках полного цикла разработки модели машинного обучения;

- для специализированных задач использование неподготовленных аннотаторов невозможно, потребуется нанимать специалистов в данной области, обладающих экспертными знаниями, что приведет к существенному удорожанию процедуры;

- современным подходом для осуществления разметки с необходимой точностью является перекрестная разметка [Ахметвалеев и др., 2021]. Но т.к. разметчики имеют разный уровень образования и опыта, то критерии разметки и присваиваемые метки могут быть несогласованными. Необходимость исправления возникающей несогласованности приводит к дополнительным затратам;

- «человеческий фактор» по определению обуславливает наличие ошибок при выполнении рутинных процедур с большим объемом данных.

Задача разметки образов на изображениях подразделяется на две подзадачи: выделения области интереса; классификации объекта в области интереса. Эти две подзадачи могут решаться как последовательно, так и совместно, в зависимости от применяемого метода.

На основании наших эмпирических данных, в ручном режиме пользователю требуется около 10 секунд на отрисовку ограничивающего прямоугольника вокруг объекта и выбор класса объекта из списка. В таком случае для разметки стандартного набора данных из 100 тысяч изображений с пятью объектами на каждое изображение потребуется около 1400 человеко-часов, что эквивалентно затратам равным примерно 600 тысяч рублей только на саму разметку данных. Добавление уровня контроля качества для ручной проверки каждого элемента размеченных данных увеличивает время подготовки выборки ещё примерно на 10% при условии выполнения контроля обученным пользователем, которому требуется примерно одна секунда на проверку каждой аннотации. При проведении контроля качества на основе перекрестной разметки количество потраченных ресурсов пропорционально количеству пользователей, работающих над пересекающимися задачами для достижения консенсуса.

Самыми затратными этапами разметки данных являются: разметка данных и проверка для контроля качества. Следовательно, важнейшая задача технологии автоматической разметки – снижение времени разметки данных и проверки результатов.

Структура решения для формирования выборок изображений естественных сцен для обучения и тестирования нейронных сетей

Функционально-логическая последовательность выполнения модулей предлагаемого решения представлена на рис. 1.

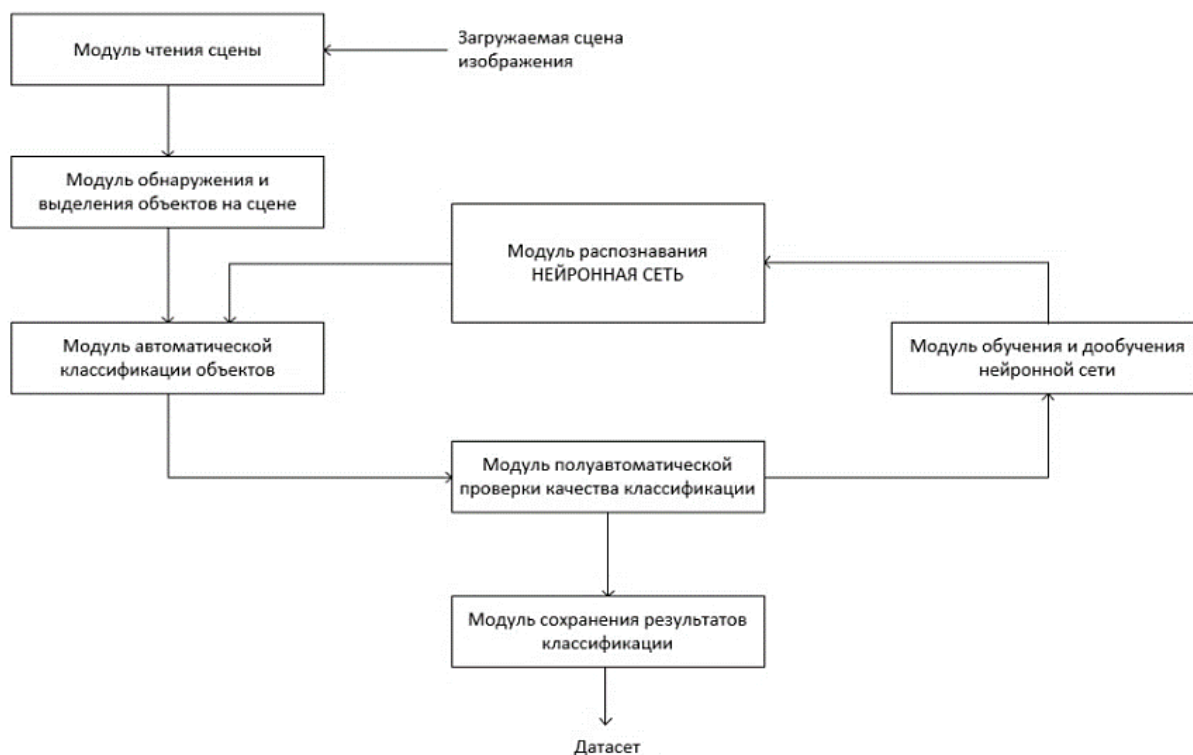


Рис. 1. Функционально-логическая последовательность выполнения модулей решения
Fig. 1. Functional-logical sequence of solution modules execution

Модуль чтения сцены отвечает за аспекты деятельности системы, не связанные непосредственно с нейронными сетями и их взаимодействием с изображениями: формальная организация изображений по проектам и датасетам, различные способы визуализации изображений и их последующая фильтрация по различным признакам, а также подготовка изображений к подаче в модули локализации и классификации.

Модуль обнаружения и выделения объектов на сцене с помощью нейронных сетей обнаруживает/сегментирует объекты на изображении и возвращает полученную информацию в модуль чтения сцены для изменения конкретного датасета изображений.

Модуль автоматической классификации объектов с помощью нейронных сетей определяет класс объекта/сущности на изображении, в случае отсутствия типов объектов, необходимых пользователю, позволяет через визуализацию результатов кластеризации присваивать классы новых типов обнаруженным объектам и возвращает полученную информацию в модуль чтения сцены для изменения конкретного датасета изображений.

Модуль полуавтоматической проверки качества классификации получает набор отдельных объектов/сущностей на изображениях вместе с классом объекта и через визуализацию позволяет скорректировать/удалить класс объекта, после чего возвращает обновленную информацию для коррекции датасета изображений.

Модуль распознавания «Нейронная сеть» создает объекты «нейронная сеть» для передачи в модули локализации и классификации, управляет реестром нейронных сетей.

Модуль обучения и дообучения нейронной сети при появлении значительного количества размеченных данных производит оптимальное (через поиск гиперпараметров) дообучение нейронных сетей на данных заказчика и возвращает полученную модель в реестр нейронных сетей.

Нейронные сети оперируют числами с плавающей точкой. Для их эффективного обучения рекомендуется использовать нормально распределенные данные, поэтому перед от-

правлением на вход нейронной сети их предварительно обрабатывают, нормализуют. Способ нормализации фактически является частью модели: при выполнении предсказаний нужно использовать тот же способ, что и при обучении.

С инженерной точки зрения при обучении нейронных сетей значительно проще не изменять размер изображений в процессе обучения: как правило, ограничивающим фактором при обучении моделей является объем памяти вычислительных устройств, поэтому уменьшение размера входных изображений ниже оптимального приводит к снижению эффективности или требует постоянного изменения размера пакета изображений на каждом шаге обучения. Широко распространенные сверточные нейронные сети испытывают значительное снижение качества предсказаний при работе с изображениями размера, сильно отличающегося от использовавшегося при обучении.

Изображения в цифровых форматах чаще всего представлены в виде трехмерной прямоугольной матрицы (два пространственных измерения и еще одно для представления трех стандартных каналов цветности RGB) целых чисел в диапазоне $[0, 255]$ (8 бит), реже в диапазоне $[0, 65536]$ (16 бит).

В результате быстрого развития технологий компьютерного зрения, изображения естественных сцен не требуют предварительной обработки. А также анализ лучших практик применения алгоритмов компьютерного зрения показал высокую эффективность решений без ручного выделения признаков, так как алгоритмы самостоятельно выделяют признаки и скрытые закономерности между ними. Поэтому было определено решение, не производить ручную отбор признаков, хранить полученные данные без изменений и проводить нормализацию прямо перед отправкой в модель.

Предварительные исследования проводились на изображениях из естественного домена, поэтому отбор признаков не производился.

В силу вышеописанных ограничений система работает без адаптации с изображениями со стороной 1200-1300 точек (пикселей) и в формате RGB 8 бит.

Новизна предлагаемого решения заключается в использовании алгоритмов компьютерного зрения не для распознавания, а для обнаружения, выделения и обработки конкретных объектов на сложных зашумленных сценах, требуемых для формирования датасетов различного назначения. Алгоритм обеспечивает автоматическое присвоение каждому распознанному объекту соответствующего тэга, используемый в дальнейшем для формирования и наполнения класса объектами со схожими признаками. В процессе обработки большого количества сцен происходит самообучение искусственного интеллекта на базе нейронной сети. В результате на выходе алгоритма формируются датасеты со структурированными и нормализованными изображениями, скомпонованными по классам, применяемых для создания обучающих и тестовых выборок, используемых для обучения нейронных сетей.

Выбор алгоритма для модуля чтения сцены и модуля обнаружения и выделения объектов на сцене

Современный уровень развития методологии применения нейронных сетей позволяет решать одновременно задачи выделения множества объектов на изображении и их классификации [Андриянов и др., 2021]. Соответственно, необходимо произвести выбор нейронной сети для построения алгоритма обнаружения объектов. Данный выбор был проведен на основе изучения сравнительных результатов испытаний различных нейронных сетей, полученных различными исследователями (например, [Сирота и др., 2019; Андриянов и др., 2021; Вожегова, 2023]), а также на собственном опыте.

Список архитектур нейронных сетей, которые были рассмотрены:

- YOLOv5 [Брехт, Коншина, 2022];
- Swin-L [Liu Z et al., 2021];
- CenterNet2 [Zhou X et al., 2021];
- YOLOv4-P7 [Wang C et al., 2021];

- Unet++ [Zhou et al., 2018];
- EfficientDet [Tan et al., 2019];
- Cascade Mask R-CNN [Liu Y et al., 2020];
- RetinaNet [Du X et al., 2020];
- Mask R-CNN [Wang J et al., 2020];
- Cascade R-CNN-FPN [Fang H et al., 2019];
- Faster R-CNN [Gao Z et al., 2019];
- Fast R-CNN [Vu T et al., 2019];
- SSD512 [Liu W et al., 2016];
- YOLO v2 + Darknet-19 [Redmon J, et al., 2017];
- Resnet-50-backbone [He et al., 2015];
- Xception-backbone [Zhang et al., 2020].

После проведения исследования и анализа различных алгоритмов для обнаружения объектов в изображениях мы пришли к выводу, что наиболее подходящим алгоритмом является YOLOv5. Этот алгоритм имеет высокую точность и скорость обнаружения объектов, а также способен работать с изображениями различных размеров и разрешений.

По результатам проведенного тестирования YOLOv5 на нашей базе данных изображений были получены следующие результаты: точность обнаружения объектов составляет около 98%, а скорость обработки одного изображения составляет менее 0,1 секунды. Это позволяет использовать YOLOv5 для обработки изображений в приложениях реального времени, в том числе таких как автономные транспортные средства, роботы и системы безопасности.

Обучение и тестирование модели на базе алгоритма YOLOv5 для модуля чтения сцены и модуля обнаружения и выделения объектов на сцене

Для модуля чтения сцены был использован датасет COCO, который содержит более 20000 изображений с аннотациями. На основе датасета была создана модель, которая способна обнаруживать объекты на изображении и классифицировать их.

В процессе обучения модели были использованы различные методы оптимизации, такие как адаптивная оценка момента (Adam) и стохастический градиентный спуск (Stochastic Gradient Descent, SGD), а также различные функции потерь, такие как перекрестная энтропия (Кросс-энтропия, Cross-Entropy), байесовская сглаженная L_1 функция потерь (Bayesian Smooth L_1 Loss) и бинарная перекрестная энтропия (Binary Cross Entropy, BCE) [Shen Yi, 2005; Ханжина, 2021].

После обучения модели она была протестирована на различных изображениях, содержащих различные объекты. Результаты показали, что модель способна эффективно обнаруживать и классифицировать объекты на изображениях.

Для оценки производительности модели на валидационном наборе данных были проведены следующие шаги:

- разделение набора данных на обучающую и валидационную выборки;
- обучение модели с использованием обучающей выборки;
- оценка производительности модели на валидационной выборке с помощью метрик, таких как точность, полнота и F-мера.

Были получены следующие результаты.

Точность модели составила 95%, что является хорошим результатом. На рис. 2 отражены результаты испытаний по определению Precision.

Полнота модели составила 80%, что также является хорошим результатом, учитывая, что модель не была обучена на полном наборе данных. На рис. 3 отражены результаты испытаний по определению Recall.

F-мера модели составила 0,87, что свидетельствует о хорошей точности и полноте модели. На рис. 4 отражены результаты испытаний по F_1 .

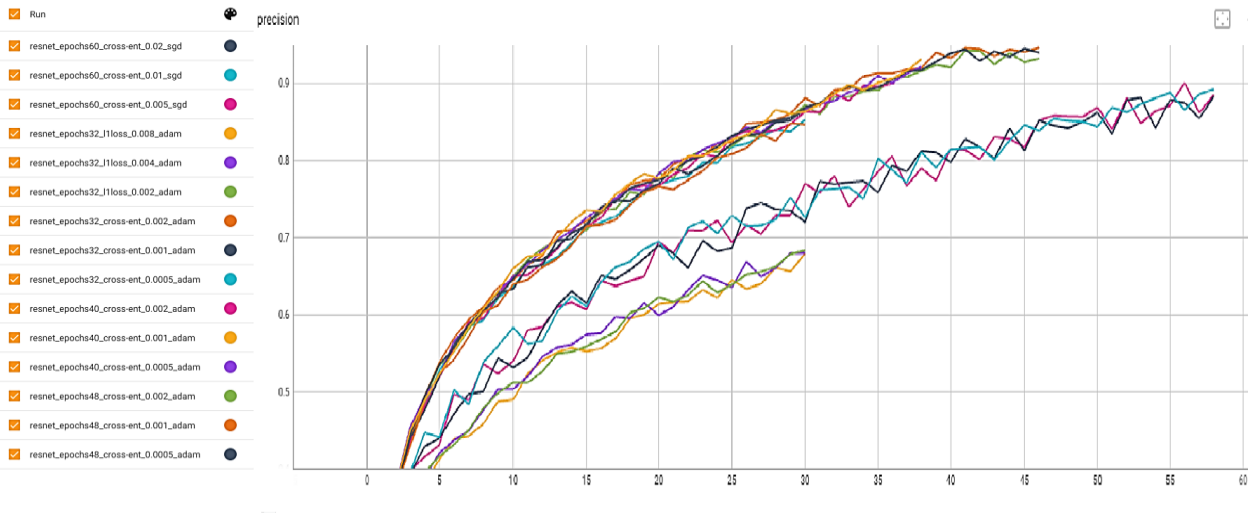


Рис. 2. Результаты испытаний: Точность
Fig. 2. Test Results: Accuracy

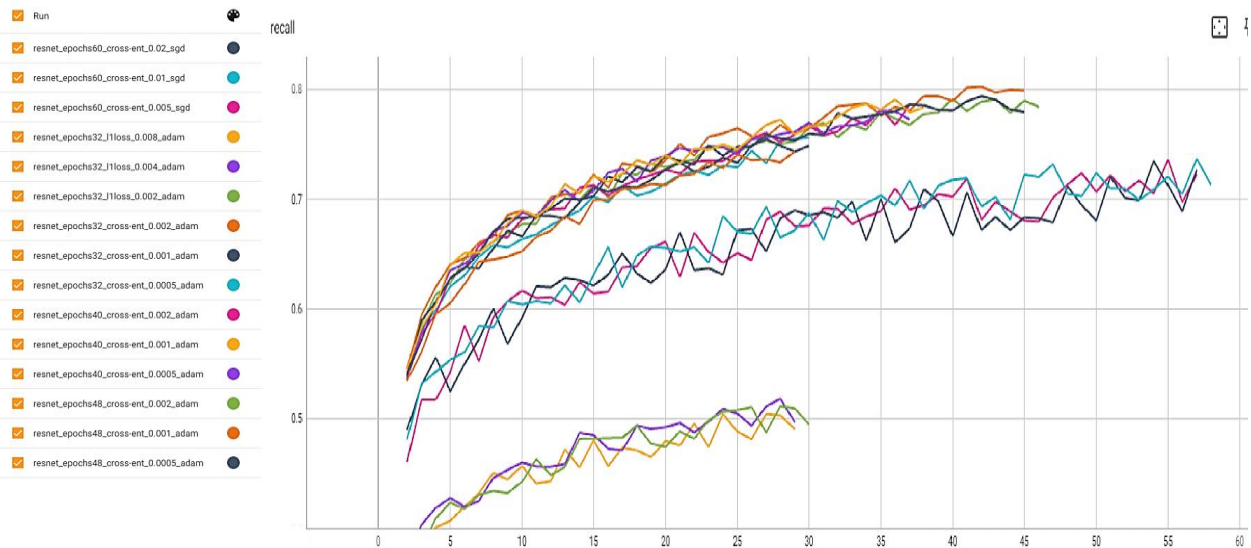


Рис. 3. Результаты испытаний: Полнота
Fig. 3. Test Results: Completeness

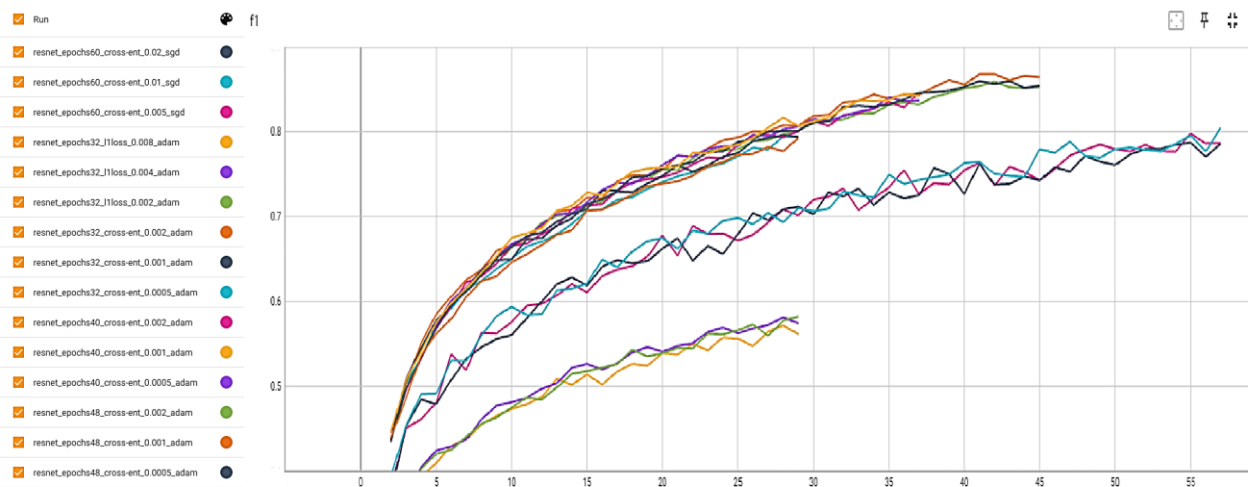


Рис. 4. Результаты испытаний: F₁-мера
Fig. 4. Test results: F₁-measure

На основе результатов оценки производительности модели можно сделать следующие выводы:

- функция потерь L_1 loss сильно уступает Cross-Entropy;
- оптимизатор SGD требует больше эпох до сходимости, но все равно заметно уступает Adam;
- сходимость происходит примерно на 40 эпохе;
- скорость обучения не влияет заметно на результат, хотя 0,001 дает немного лучшие результаты.

Таким образом, можно сделать вывод, что модель успешно справляется с задачей обнаружения и выделения объектов на сцене с высокой точностью и полнотой, а также демонстрирует хороший баланс между точностью и полнотой и может быть использована для дальнейшей работы в реальных условиях.

Заключение

В ходе выполнения работы предлагаемое решение достигает следующих параметров.

Ускорение времени разметки данных прямоугольниками (для задачи детекции) не менее чем в 50 раз по сравнению с ручной разметкой, осуществляемой квалифицированным разметчиком в условиях нормирования, хронометража рабочего времени, качества исходных данных и достаточном объеме набора данных. При оценке было принято, что в зависимости от сложности сцены среднее время разметки прямоугольниками в ручном режиме составляет 200-250 объектов в час на одного разметчика.

Снижение стоимости сегментирования и аннотирования данных в задачах сегментации не менее чем в 2 раза по сравнению с ручной сегментацией, осуществляемой квалифицированным разметчиком.

Уменьшение количества ошибочно размеченных данных не менее чем на 20% по сравнению с результатами ручной разметки в аналогичных условиях. Параметр проверялся путем верификации тестовых наборов данных, полученных в ручном режиме, и полученных с помощью предлагаемого решения, и соотношением в количественном выражении выявленных ошибочных данных в каждом использованном методе.

Улучшение качества сегментирования данных не менее чем на 10% по сравнению с ручной сегментацией в аналогичных условиях. Параметр проверяется с помощью показателей метрики IoU (Intersection-over-Union) на размеченных данных, полученных в ручном режиме и полученных с помощью предлагаемого решения.

Список литературы

- Ахметвалеев Р.Р., Шабанова К.И., Падукова А.А., Лакман И.А. 2021. Методика разметки медицинских изображений с функцией кросс-проверки и интеллектуального сегментирования. Вестник ВШОУЗ, 7, 3: 62-69.
- Андрянов Н.А. 2022. Обнаружение объектов на изображении: от критериев Байеса и Неймана–Пирсона к детекторам на базе нейронных сетей EfficientDet. Компьютерная оптика. 46, 1: 139-159. DOI: 10.18287/2412-6179-CO-922.
- Брехт Э.А., Коншина В.Н. 2022. Применение нейронной сети YOLO для распознавания дефектов. Intellectual Technologies on Transport, 2: 41-47. DOI: 10.24412/2413-2527-2022-230-41-47.
- Вожегова М.А. 2023. Структурно-функциональная модель комплексов неразрушающего контроля с системой обработки данных на основе нейронных сетей. Экономика. Информатика. 50(2): 389–397. DOI: 10.52575/2687-0932-2023-50-2-389-397.
- Сирота А.А., Митрофанова Е.Ю., Милованова А.И. 2019. Анализ алгоритмов поиска объектов на изображениях с использованием различных модификаций сверточных нейронных сетей. Вестник ВГУ, серия: системный анализ и информационные технологии, 3: 123-137.
- Ханжина Н.Е. 2021. Байесовские функции потерь для моделирования гомоскедастичной алеаторной неопределенности в задаче детекции пылицы на изображениях. Научно-технический

- вестник информационных технологий, механики и оптики. 21, 4: 535–544. DOI: 10.17586/2226-1494-2021-21-4-535-544.
- Du X, Lin T, Jin P, Ghiasi G, Tan M, Cui Y, Le QV, Song X. 2020. SpineNet: Learning scale-permuted backbone for recognition and localization. Proc IEEE Conf on Computer Vision and Pattern Recognition (CVPR), 1: 11593-11601. DOI: 10.1109/CVPR42600.2020.01161.
- Fang H, Sun J, Wang R, Gou M, Li Y, Lu C, Tong SJ. 2019. InstaBoost: Boosting instance segmentation via probability map guided copy-pasting. Proc 2019 IEEE/CVF Int Conf on Computer Vision (ICCV), 1: 682-691. DOI: 10.1109/ICCV.2019.00077.
- Gao Z, Wang L, Wu G. 2019. LIP: Local importance-based pooling. Proc 2019 IEEE/CVF Int Conf on Computer Vision (ICCV), 1: 3355-3364. DOI: 10.1109/ICCV.2019.00345.
- Hsu G., Chen J., Chung Y. 2013. Application-oriented license plate recognition. IEEE Trans. Veh. Technol. 62(2): 552-561.
- Liu Y, Wang S, Liang T, Zhao Q, Tang Z, Ling H. CBNet: A novel composite backbone network architecture for object detection. arXiv Preprint. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/6834>.
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C, Berg A. 2016. SSD: Single shot multibox detector. Proc European Conf on Computer Vision (ECCV), 1: 1-17. DOI: 10.1007/978-3-319-46448-0_2.
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Zh, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. arXiv Preprint. URL: <https://arxiv.org/pdf/2103.14030v1.pdf>.
- Redmon J, Farhadi A. 2017. YOLO9000: Better, faster, stronger. Proc 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR), 1: 7263-7271. DOI: 10.1109/CVPR.2017.690.
- Shen Yi. 2005. Loss Functions For Binary Classification and Class Probability Estimation. University of Pennsylvania. URL: <http://stat.wharton.upenn.edu/~buja/PAPERS/yi-shen-dissertation.pdf>.
- Tan M, Pang R, Le QV. EfficientDet: Scalable and efficient object detection. arXiv Preprint. URL: <https://arxiv.org/abs/1911.09070>.
- Vu T, Jang H, Pham T, Yoo C. 2019. Cascade RPN: Delving into high-quality region proposal network with adaptive convolution. Proc 33rd Conf on Neural Information Processing Systems (NeurIPS 2019), 1: 1-11.
- Wang C, Bochkovskiy A, Liao H. Scaled-YOLOv4: Scaling cross stage partial network. arXiv Preprint. URL: <https://arxiv.org/pdf/2011.08036v2.pdf>.
- Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, Liu W, Xiao B. 2020. Deep highresolution representation learning for visual recognition. IEEE Trans Pattern Anal Mach Intell, 1: 1-23. DOI: 10.1109/tpami.2020.2983686.
- Zhang R, Du L, Xiao Q, Liu J. 2020. Comparison of Backbones for Semantic Segmentation Network. J. Phys.: Conf. Ser. 1544 012196. DOI: 10.1088/1742-6596/1544/1/012196.
- Zhou X, Koltun V, Krahenbuhl P. Probabilistic two-stage detection. arXiv Preprint. URL: <https://arxiv.org/pdf/2103.07461v1.pdf>.

References

- 1Ahmetvaleev R.R., Shabanova K.I., Padukova A.A., Lakman I.A. 2021. Methodology of marking medical images with the function of cross-check and intellectual segmentation. Vestnik VSHOUZ, 7, 3: 62-69. (in Russian)
- Andriyanov N.A., Dementiev V.E., Tashlinskii A.G. 2022. Detection of objects in the images: from likelihood relationships towards scalable and efficient neural networks. Computer Optics, 46(1): 139-159. DOI: 10.18287/2412-6179-CO-922. (in Russian)
- Brekht E.A., Konshina V.N. 2022. Application of YOLO Neural Network for Defect Recognition. Intellectual Technologies on Transport, 2: 41-47. DOI: 10.24412/2413-2527-2022-230-41-47 (data access: 28 march 2023). (in Russian)
- Vozhegova M.A. 2023. Structural and Functional Model of Non-Destructive Testing Complexes with a Data Processing System Based on Neural Networks. Economics. Information Technologies, 50(2): 389–397. DOI: 10.52575/2687-0932-2023-50-2-389-397 (in Russian)
- Sirota A.A., Mitrofanova E.Yu., Milovanova A.I. 2019. Analysis of algorithms for searching objects in images using various modifications of convolutional neural network. Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies, 3: 123-137. (in Russian)

- Khanzhina N.E. 2021. Bayesian losses for homoscedastic aleatoric uncertainty modeling in pollen image detection. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 21, 4: 535–544. DOI: 10.17586/2226-1494-2021-21-4-535-544. (in Russian)
- Du X, Lin T, Jin P, Ghiasi G, Tan M, Cui Y, Le QV, Song X. 2020. SpineNet: Learning scale-permuted backbone for recognition and localization. *Proc IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*, 1: 11593-11601. DOI: 10.1109/CVPR42600.2020.01161.
- Fang H, Sun J, Wang R, Gou M, Li Y, Lu C, Tong SJ. 2019. InstaBoost: Boosting instance segmentation via probability map guided copy-pasting. *Proc 2019 IEEE/CVF Int Conf on Computer Vision (ICCV)*, 1: 682-691. DOI: 10.1109/ICCV.2019.00077.
- Gao Z, Wang L, Wu G. 2019. LIP: Local importance-based pooling. *Proc 2019 IEEE/CVF Int Conf on Computer Vision (ICCV)*, 1: 3355-3364. DOI: 10.1109/ICCV.2019.00345.
- Hsu G., Chen J., Chung Y. 2013. Application-oriented license plate recognition. *IEEE Trans. Veh. Technol.* 62(2): 552-561.
- Liu Y, Wang S, Liang T, Zhao Q, Tang Z, Ling H. CBNet: A novel composite backbone network architecture for object detection. *arXiv Preprint*. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/6834>.
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C, Berg A. 2016. SSD: Single shot multibox detector. *Proc European Conf on Computer Vision (ECCV)*, 1: 1-17. DOI: 10.1007/978-3-319-46448-0_2.
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Zh, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv Preprint*. URL: <https://arxiv.org/pdf/2103.14030v1.pdf>.
- Redmon J, Farhadi A. 2017. YOLO9000: Better, faster, stronger. *Proc 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*, 1: 7263-7271. DOI: 10.1109/CVPR.2017.690.
- Shen Yi. 2005. Loss Functions For Binary Classification and Class Probability Estimation. University of Pennsylvania. URL: <http://stat.wharton.upenn.edu/~buja/PAPERS/yi-shen-dissertation.pdf>.
- Tan M, Pang R, Le QV. EfficientDet: Scalable and efficient object detection. *arXiv Preprint*. URL: <https://arxiv.org/abs/1911.09070>.
- Vu T, Jang H, Pham T, Yoo C. 2019. Cascade RPN: Delving into high-quality region proposal network with adaptive convolution. *Proc 33rd Conf on Neural Information Processing Systems (NeurIPS 2019)*, 1: 1-11.
- Wang C, Bochkovskiy A, Liao H. Scaled-YOLOv4: Scaling cross stage partial network. *arXiv Preprint*. URL: <https://arxiv.org/pdf/2011.08036v2.pdf>.
- Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, Liu W, Xiao B. 2020. Deep highresolution representation learning for visual recognition. *IEEE Trans Pattern Anal Mach Intell*, 1: 1-23. DOI: 10.1109/tpami.2020.2983686.
- Zhang R, Du L, Xiao Q, Liu J. 2020. Comparison of Backbones for Semantic Segmentation Network. *J. Phys.: Conf. Ser.* 1544 012196. DOI: 10.1088/1742-6596/1544/1/012196.
- Zhou X, Koltun V, Krahenbuhl P. Probabilistic two-stage detection. *arXiv Preprint*. URL: <https://arxiv.org/pdf/2103.07461v1.pdf>.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

Conflict of interest: о potential conflict of interest related to this article was reported.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Гусев Александр Валерьевич, Научный руководитель проектов. ООО «Техинтегратор», г. Москва, Россия

INFORMATION ABOUT THE AUTHOR

Aleksandr V. Gusev, Scientific Project Manager. Limited Liability Company «Techintegrator», Moscow, Russian Federation