

УДК 004.89

DOI 10.52928/2070-1624-2023-41-2-29-34

**МЕТОД АУГМЕНТАЦИИ ДАННЫХ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА РАСПОЗНАВАНИЯ
ИЗОБРАЖЕНИЙ ЭЛЕКТРОННОЙ КОММЕРЦИИ****В. В. СОРОКИНА¹,***акад. НАН Беларуси, д-р техн. наук, проф. С. В. АБЛАМЕЙКО^{1, 2}*¹*(Белорусский государственный университет, Минск)*²*(Объединённый институт проблем информатики НАН Беларуси, Минск)*

В быстро развивающейся сфере электронной коммерции визуальное представление товаров играет ключевую роль в привлечении потребителей и повышении коэффициента конверсии. В статье представлен новый метод аугментации изображений, который включает в себя сегментацию изображений, определение доминирующего цвета, замену фона и реалистичную генерацию теней. В совокупности эти шаги способствуют созданию аугментированных изображений, которые не только используются в электронных каталогах, но и расширяют возможности нейронных сетей разнообразными и обогащенными обучающими данными. Построенный метод позволяет решить проблемы, связанные с классовым дисбалансом, усилить обобщения модели, а также улучшить качество распознавания.

Ключевые слова: *сверточная нейронная сеть, распознавание изображений, сегментация, генерация теней, электронная коммерция.*

Введение. Аугментация данных играет ключевую роль в различных областях машинного обучения, компьютерного зрения и искусственного интеллекта. Ее важность заключается в способности повышать качество и количество обучающих данных, что, в свою очередь, способствует повышению производительности и надежности моделей машинного обучения. Важность аугментации данных обусловлена следующими причинами:

- широкое разнообразие: методы аугментации данных вносят разнообразие в набор обучающих данных за счет создания новых, измененных версий существующих данных. Такое разнообразие помогает моделям машинного и глубокого обучения лучше обобщать невидимые данные и снижает риск переобучения. Модели, обученные на более разнообразных данных, зачастую надежнее и с меньшей вероятностью дают неправильные прогнозы;

- улучшенная генерализация модели: аугментация данных обеспечивает более широкое представление основного распределения данных, позволяя моделям охватывать более широкий спектр закономерностей и вариаций. В результате создаются модели, которые лучше адаптированы к реальным сценариям, где данные могут сильно различаться;

- решение проблемы нехватки данных: во многих реальных приложениях, особенно в сфере электронной коммерции, получение большого набора размеченных данных может оказаться сложной и дорогостоящей задачей. Аугментация данных позволяет специалистам искусственно увеличивать размер своего набора данных, что делает более целесообразным обучение сложных моделей, даже если данных недостаточно;

- уменьшение систематической ошибки: аугментация данных может уменьшить систематическую ошибку в обучающих данных. Путем расширения выборок для недостаточно представленных классов можно сбалансировать набор данных и гарантировать, что модель не отдаст предпочтение классу большинства, что приводит к более справедливым и точным прогнозам;

- сохранение конфиденциальности: в ситуациях, когда существует ограничение доступа к конфиденциальным данным, можно использовать аугментацию данных для создания синтетических данных, которые сохраняют статистические свойства исходных данных, одновременно защищая оригинальные данные;

- регуляризация: аугментация служит формой регуляризации, не позволяя моделям запоминать обучающие данные, представляя их в различных вариациях. Это помогает предотвратить переобучение, когда модель хорошо работает на обучающих, но плохо на тестовых данных;

- трансферное обучение: аугментация имеет важное значение в сценариях трансферного обучения. Предварительно обученные модели можно точно настроить для новой задачи с помощью аугментированных данных, что позволяет им адаптироваться к конкретным нюансам новой проблемы, извлекая при этом пользу из знаний, полученных из исходных данных;

- реалистичное тестирование: аугментация также можно использовать для создания расширенных тестовых данных, что важно для оценки надежности и производительности модели в реальных сценариях, особенно в сфере электронной коммерции, где количество продуктов постоянно увеличивается.

Решение проблемы нехватки данных в сфере электронной коммерции позволяет малому бизнесу развиваться быстрее и быть более конкурентоспособным. С научной точки зрения аугментация открывает широкий спектр возможных применений.

В данной статье демонстрируется новый метод аугментации изображений электронной коммерции: аугментация производится за счет сегментации изображения, замены фона, генерации тени и добавления логотипа. Построенный метод аугментации подходит для задач распознавания изображений электронной коммерции при создании каталога товаров.

Суть этой инновации заключается в разработке интеллектуальной системы аугментации данных, предназначенной для повышения производительности нейронной сети за счет обогащения ее набора обучающих данных. Возможности этой системы выходят за рамки традиционных подходов к аугментации, поскольку она позволяет нейронной сети воспринимать изображения продуктов на разном фоне, в различных условиях освещения и во множестве контекстных настроек.

В работе предлагается метод аугментации данных для распознавания изображений товаров электронной коммерции на основе сверточных нейронных сетей. На основе данного метода создана система, которая позволяет пользователю создавать аугментированные изображения путем изменения фона изображений, а также добавлять к изображениям различные тени и наклейки или надписи. Система способна готовить изображения для задач распознавания изображений, таких как обнаружение и сегментация изображений в электронной коммерции. Среднее время вычислений для генерации одного кадра составляет 1,2 с.

Анализ существующих подходов. Визуальная привлекательность изображений товаров электронной коммерции играет важную роль в привлечении и удержании клиентов. В нескольких исследованиях для улучшения изображений продуктов использовались методы аугментации изображений. Классическая аугментация включает горизонтальное и вертикальное вращение, масштабирование и настройку цвета [1]. Такие методы не только улучшают визуальную эстетику изображений продуктов, но и дополняют набор обучающих данных, в результате чего модели становятся более устойчивыми к изменениям освещения, ориентации и фона. Однако традиционные методы увеличения данных недостаточны в сфере электронной коммерции из-за ее специфики.

Для сферы электронной коммерции характерны две основные проблемы, связанные с данными: проблема дисбаланса классов, обычно называемая «проблемой длинного хвоста» и проблема корректной аугментации данных для подготовки электронного каталога. В зависимости от типа проблемы существуют различные техники ее устранения. Так, например, для решения проблемы дисбаланса классов в [2] был предложен метод, который динамически балансирует положительные и отрицательные градиенты выборки для каждой категории, уменьшая дисбаланс классов. Однако для его оптимальной производительности нужна точная настройка гиперпараметров. Также алгоритм не может полностью устранить дисбаланс классов в крайне несбалансированных наборах данных. Предложенный в [3] метод эффективно снижает конкуренцию между редкими и частыми категориями, группируя их в непересекающиеся наборы, но может работать неоптимально, если категории естественным образом не делятся на четко определенные группы. Исследования [4; 5] предлагают техники сосредоточивания на редких категориях, при этом требуя точной настройки параметров, не гарантируя совместимость с определенными архитектурами нейронных сетей. Для решения задачи аугментации данных используются техники MixUp [6], CopyPaste [7] и CutOut [8]: MixUp увеличивает разнообразие наборов данных за счет смешивания двух изображений, уменьшая переобучение, но может привести к появлению артефактов на изображениях; CopyPaste добавляет разнообразие, случайным образом вставляя логотипы или элементы с одного изображения на другое, однако необходим тщательный контроль процесса вставки, чтобы избежать нестыковок; CutOut вводит окклюзии и побуждает модель учиться на частичной информации, при этом может удалить важную информацию с изображений.

Подводя итог, следует отметить, что представленные методы обладают определенными преимуществами в устранении классового дисбаланса и увеличении разнообразия наборов данных. Однако их эффективность может зависеть от конкретного набора данных и решаемой задачи. Также для достижения оптимальных результатов может потребоваться тщательная настройка параметров. Кроме того, некоторые методы могут потребовать компромисса между улучшенным балансом и потенциальными недостатками, такими как увеличение времени обучения или сложность модели.

Предложенный нами метод позволяет генерировать аугментированные данные непосредственно для сферы электронной коммерции (метод позволяет автоматически создать новое изображение на основе оригинального при помощи изменения фона, добавления теней и логотипов), учитывая ее специфику, при этом соблюдая баланс времени обучения и сложности модели, что выделяет его среди существующих подходов.

Метод. В ходе исследования был разработан метод, позволяющий автоматически генерировать аугментированные изображения без привлечения человеческих ресурсов. Он включает в себя следующие компоненты:

- модуль сегментации;
- модуль определения доминирующего цвета;
- модуль удаления и замены фона;
- модуль генерации теней;
- модуль добавления наклейки/надписи (логотипа).

Схема построенного метода аугментации показана на рисунке 1.

Рисунок 1. – Схема метода аугментации



На вход алгоритма поступает оригинальное RGB-изображение. На первом этапе происходит сегментация изображения, выделенный объект далее передается на вход модуля определения доминантного цвета и модуля замены фона. Определение доминантного цвета необходимо, чтобы при замене фона избежать ситуации, когда объект полностью с ним сливается. Затем происходит генерация одного из трех типов тени: зеркальной, падающей или реалистичной. В конце на полученное изображение добавляется логотип.

Модуль сегментации. В начале производится сегментация изображения, в результате чего выделяются передний и задний планы. За это отвечает реализованный нами алгоритм на основе нейронной сети YOLACT и стандартизации весов, подробное описание которого представлено в [9]. Обучение модели было направлено на распознавание 21 класса объектов электронной коммерции. Стандартизация весов использовалась в сверточных слоях при прямом проходе обучения нейронной сети.

Модуль определения доминирующего цвета. После получения сегментированного изображения для полученного объекта определяется его доминантный цвет. Для этого используется метод k-means – алгоритм обучения без учителя для кластеризации данных в K кластеров. Идея метода заключается в минимизации суммарного квадратичного отклонения точек кластеров от центров этих кластеров. Конечным результатом является определение K доминирующих цветов в изображении, представленных K центроидами кластера. При подборе правильного значения K центр тяжести самого большого кластера будет довольно хорошим представлением доминирующего цвета изображения.

Модуль удаления и замены фона. Данный модуль представляет собой сочетание 2 методов: метода удаления фона и метода его замены.

Удаление фона происходит с помощью альфа-смешивания: объект переднего плана совмещается с прозрачным фоном на основе альфа-маски. Формула альфа-смешивания объединяет цвета изображения переднего плана (F) и фона (B) на основе их значений альфа.

Результирующий цвет (C) рассчитывается по формуле

$$C = \alpha * F + (1 - \alpha) * B.$$

Система предлагает 3 варианта замены фона:

– сплошная заливка: замена фона сплошным цветом или градиентом. Цвет нового фона выбирается случайным образом, однако не совпадает с доминантным цветом сегментированного объекта;

– замена на основе текстуры: наложение сегментированного объекта на определенную текстуру. В системе представлено более 30 различных текстур;

– композиция изображения: объединение объекта переднего плана с новым фоном путем альфа-смешивания.

Модуль генерации теней. Представляет собой алгоритм, позволяющий генерировать зеркальную, падающую и реалистичную тени. Алгоритмы зеркальной и падающей теней реализованы при помощи методов компьютерного зрения, позволяющих придавать теням мягкость, моделируя полутень (область, где тень переходит от полностью освещенной к полностью затененной), а также использующих различные методы сглаживания, чтобы уменьшить неровные края и улучшить общее качество теней и отражений.

Метод генерации реалистичной тени основан на нашем исследовании [10], где предлагается новый подход для генерации двумерных теней в сфере электронной коммерции. Идея метода – при использовании глубокой нейронной сети, модифицированной моделью трансформера, извлечь структурную информацию из изображений и преобразовать карты признаков нижнего уровня. Представленный алгоритм показал при тестировании улучшение точности и качества генерируемых теней в сравнении с существующими подходами.

Модуль добавления наклейки/надписи (логотипа). Данный модуль принимает на вход RGB-изображение с замененным фоном и сгенерированной тенью, логотип (надпись), а также координаты расположения объекта внутри фона на основе его сегментации. Далее происходит совмещение изображения с логотипом таким образом, что он не перекрывает основной объект, в результате чего формируется высококачественное сегментированное изображение.

Результаты. В ходе исследования была разработана система аугментации данных, позволяющая автоматически генерировать изображения, удовлетворяющие критериям сферы электронной коммерции. Пример аугментированного изображения представлен на рисунке 2. Для данного примера в настройках системы были выбраны изображение утюга на сером градиентном фоне, новый фон в виде сцены со светом, логотип и эффект зеркальной тени. Алгоритм автоматически произвел сегментацию изображения и замену фона, сгенерировал зеркальную тень и добавил логотип, не перекрывая сам объект.

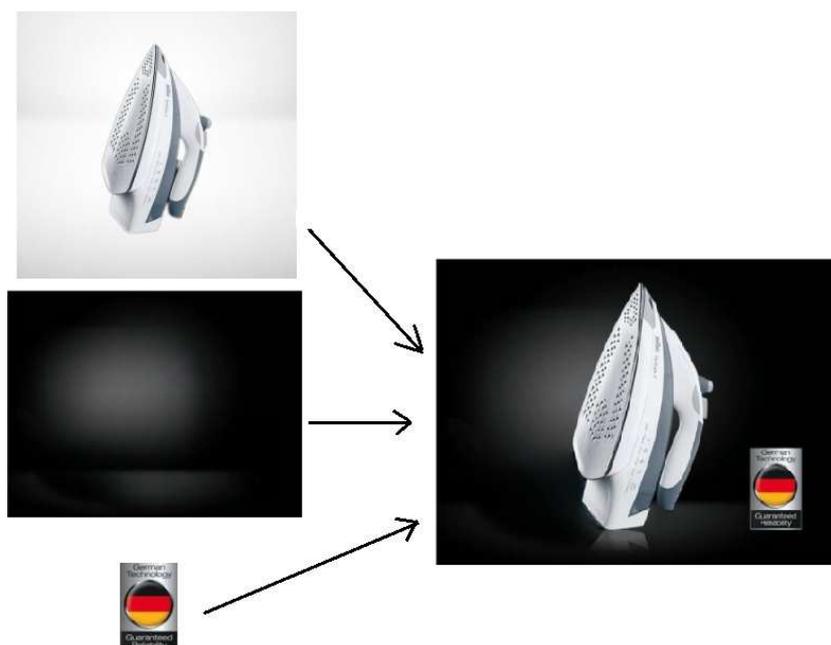


Рисунок 2. – Аугментированное изображение

Для проверки качества аугментации система была протестирована при решении задачи сегментации с использованием простой модели на изначально небольшом наборе данных:

1. Модель: для этого исследования была выбрана архитектура U-Net [11]. U-Net хорошо подходит для задач сегментации изображений, что делает ее подходящим кандидатом для оценки эффективности аугментации данных в контексте изображений продуктов электронной коммерции.

2. Собственный набор данных [9]:

– базовый: собственный набор изображений товаров электронной коммерции, включающий 1500 изображений, а также маски сегментации на уровне пикселей для каждой категории товаров;

– аугментированный при помощи построенной системы набор данных на основе базового.

3. Оценка: обученная модель оценивается на отдельном тестовом наборе.

4. Метрики сравнения:

- точность: общая точность сегментации в тестовом наборе данных;
- IoU (пересечение над объединением): измеряет перекрытие между предсказанными и реальными масками;
- mAP (средняя точность): оценивает производительность обнаружения и сегментации объектов модели;
- время сходимости;
- время обучения.

Стратегия обучения модели: обучить U-Net при одинаковой предобработке данных на базовом и аугментированном наборах данных.

5. Добавить новый модуль сегментации в предварительно обученную модель Mask R-CNN.

Результаты:

1. Обучение на базовом наборе:

- точность: 85 %;
- IoU: 0,72;
- mAP: 0,78;
- время сходимости: 150 эпох;
- время обучения: 4 часа.

2. Обучение на аугментированном наборе:

- точность: 92 %;
- IoU: 0,85;
- mAP: 0,89;
- время сходимости: 120 эпох;
- время обучения: 5 часов.

Результаты показывают, что предложенный метод аугментации данных оказывает положительное влияние на качество распознавания изображений электронной коммерции с помощью модели U-Net, обеспечивая более высокую точность и более быструю сходимость. Компромисс в виде немного более длительного времени обучения оправдан улучшенным качеством распознавания.

Заключение. Решение таких задач, как сегментация изображений, определение основного цвета, генерация теней позволило создать технологию распознавания с помощью аугментации данных, что может быть использовано как для повышения качества обучаемых моделей и улучшения качества распознавания изображений электронной коммерции, так и для автоматической подготовки изображений для электронного каталога товаров. Данная технология была протестирована при решении задачи сегментации изображений товаров электронной коммерции и показала улучшение точности и качества модели вместе с более быстрой сходимостью.

ЛИТЕРАТУРА

1. Shorten C., Khoshgoftaar T. M. A survey on Image Data Augmentation for Deep Learning // *J. of Big Data*. – 2019. – Vol. 6:60. – DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
2. Wang, J., Zhang, W., Zang, Y. et al. Seesaw Loss for Long-Tailed Instance Segmentation // 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. – P. 9690–9699. – DOI: [10.1109/CVPR46437.2021.00957](https://doi.org/10.1109/CVPR46437.2021.00957).
3. Li Y., Wang T., Kang B. et al. Overcoming Classifier Imbalance for Long-Tail Object Detection with Balanced Group Softmax // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. – P. 10988–10997. – DOI: [10.1109/CVPR42600.2020.01100](https://doi.org/10.1109/CVPR42600.2020.01100).
4. Tan J., Lu X., Zhang G. et al. Equalization Loss v2: A New Gradient Balance Approach for Long-tailed Object Detection // 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. – P. 1685–1694. – DOI: [10.1109/CVPR46437.2021.00173](https://doi.org/10.1109/CVPR46437.2021.00173).
5. Esposito, C., Landrum, G. A., Schneider N. et al. GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning // *J. Chem. Inf. Model.* – 2021. – Vol. 61, iss. 6. – P. 2623–2640. – DOI: [10.1021/acs.jcim.1c00160](https://doi.org/10.1021/acs.jcim.1c00160).
6. Chen Y., Hu V. T., Gavves E. et al. PointMixup: Augmentation for Point Clouds // ECCV 2020. Lecture Notes in Computer Science. – 2020. – Vol. 12348. P. 330–345. – DOI: [10.1007/978-3-030-58580-8_20](https://doi.org/10.1007/978-3-030-58580-8_20).
7. Ghiasi G., Cui, Y., Srinivas, A. et al. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation // 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. – P. 2917–2927. – DOI: [10.1109/CVPR46437.2021.00294](https://doi.org/10.1109/CVPR46437.2021.00294).
8. Devries, T., Taylor, G. W. Improved Regularization of Convolutional Neural Networks with Cutout // *ArXiv*. – 2017. – Vol. 1708.04552. – DOI: [10.48550/arXiv.1708.04552](https://doi.org/10.48550/arXiv.1708.04552).
9. V. Sorokina, S. Ablameyko Neural network training acceleration by weight standardization in segmentation of electronic commerce images // *Reliability Engineering and Computational Intelligence. Studies in Computational Intelligence*. – 2021. – Vol. 976. – P. 237–244. – DOI: [10.1007/978-3-030-74556-1_14](https://doi.org/10.1007/978-3-030-74556-1_14).

10. Sorokina V., Ablameyko S. 2D Cast Shadow Generation in E-commerce Image Using UNet Vision Transformer // 2023 Intern. Conf. on Information and Digital Technologies, Zilina, Slovakia, 20–22 June 2023. – P. 33–36. – DOI: [10.1109/IDT59031.2023.10194446](https://doi.org/10.1109/IDT59031.2023.10194446).
11. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention. – 2015. – Vol. 9351. – P. 234–241. – DOI: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).

REFERENCES

1. Shorten, C., Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, (6), 60. DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
2. Wang, J., Zhang, W., Zang, Y., Cao, Y., Pang, J., Gong, T., ... Lin, D. (2021). Seesaw Loss for Long-Tailed Instance Segmentation. In *2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (9690–9699). IEEE. DOI: [10.1109/CVPR46437.2021.00957](https://doi.org/10.1109/CVPR46437.2021.00957).
3. Li, Y., Wang, T., Kang, B., Tang, S., Wang, C., Li, J., & Feng, J. (2020). Overcoming Classifier Imbalance for Long-Tail Object Detection with Balanced Group Softmax. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (10988–10997). IEEE. DOI: [10.1109/CVPR42600.2020.01100](https://doi.org/10.1109/CVPR42600.2020.01100).
4. Tan, J., Lu, X., Zhang, G., Yin, C., & Li, Q. (2020). Equalization Loss v2: A New Gradient Balance Approach for Long-tailed Object Detection. In *2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (1685–1694). IEEE. DOI: [10.1109/CVPR46437.2021.00173](https://doi.org/10.1109/CVPR46437.2021.00173).
5. Esposito, C., Landrum, G. A., Schneider N., Stiefl N., & Riniker, S. (2021). GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning. *J. Chem. Inf. Model.*, 61(6), 2623–2640. DOI: [10.1021/acs.jcim.1c00160](https://doi.org/10.1021/acs.jcim.1c00160).
6. Chen, Y., Hu, V. T., Gavves, E., Mensink, T., Mettes, P., Yang, P. & Snoek C. G. M. (2020). PointMixup: Augmentation for Point Clouds. In A. Vedaldi, H. Bischof, T. Brox, & J. M. Frahm (Eds.), *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science*, vol. 12348 (330–345). Springer, Cham. DOI: [10.1007/978-3-030-58580-8_20](https://doi.org/10.1007/978-3-030-58580-8_20).
7. Ghiasi G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., ... Zoph, B. (2021). Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. In *2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (2917–2927). IEEE. DOI: [10.1109/CVPR46437.2021.00294](https://doi.org/10.1109/CVPR46437.2021.00294).
8. Devries, T., & Taylor, G. W. (2017). Improved Regularization of Convolutional Neural Networks with Cutout. *ArXiv*, (1708.04552). DOI: [10.48550/arXiv.1708.04552](https://doi.org/10.48550/arXiv.1708.04552).
9. Sorokina, V., & Ablameyko, S. (2021). Neural Network Training Acceleration by Weight Standardization in Segmentation of Electronic Commerce Images. In C. van Gulijk, & E. Zaitseva (Eds.), *Reliability Engineering and Computational Intelligence. Studies in Computational Intelligence: Vol. 976* (237–244). Springer, Cham. DOI: [10.1007/978-3-030-74556-1_14](https://doi.org/10.1007/978-3-030-74556-1_14).
10. Sorokina, V., & Ablameyko, S. (2023). 2D Cast Shadow Generation in E-commerce Image Using UNet Vision Transformer. In *2023 International Conference on Information and Digital Technologies (IDT)* (31–36). IEEE. DOI: [10.1109/IDT59031.2023.10194446](https://doi.org/10.1109/IDT59031.2023.10194446).
11. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science. Vol. 9351* (234–241). Springer, Cham. DOI: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).

Поступила 23.10.2023

DATA AUGMENTATION METHOD TO IMPROVE THE QUALITY OF E-COMMERCE IMAGE RECOGNITION

V. SOROKINA¹, S. ABLAMEYKO^{1,2}

¹(Belarusian State University, Minsk)

²(United Institute of Informatics Problems of National Academy of Sciences of Belarus, Minsk)

In the rapidly evolving landscape of e-commerce, the visual representation of products plays a pivotal role in engaging consumers and driving conversion rates. This article introduces a new approach for image augmentation that includes objects segmentation, dominant color determination, background replacement and realistic shadow generation. These steps collectively contribute to the creation of augmented images that are used not only in the electronic catalogues but enrich abilities of the neural networks with various and fortified training data. Developed system allows to solve problems related to class imbalance and to enhance model generalization as well as to improve the quality of recognition.

Keywords: convolutional neural network, image recognition, segmentation, shadow generation, e-commerce.