

УДК 004.032.26

МЕТОДИКА ПОВЫШЕНИЯ УСТОЙЧИВОСТИ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ АУГМЕНТАЦИИ ДАННЫХ НА ПРИМЕРЕ МОДЕЛИ ДЛЯ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ ГЛАЗНЫХ ЯБЛОК



Н.А. Резников

Студент кафедры
электронной техники и
технологии БГУИР
m1ne.turtle467@gmail.com



А.В. Шкрабов

Студент кафедры электронной
техники и технологии БГУИР
sashashkrabov04@gmail.com



С. К. Дик

Заведующий кафедрой
инженерной и компьютерной
графики БГУИР,
канд. физ.-мат. наук, доцент
sdick@bsuir.by



Г.Д. Ситник

Доцент кафедры общей врачебной практики
с курсом гериатрии и паллиативной медицины
БГМУ,
канд. мед. наук, доцент, врач высшей категории
по неврологии



И.И. Ревинская

Старший преподаватель кафедры
электронной техники и технологии
БГУИР
inna_revinskaya@bsuir.by

Н.А. Резников

Студент кафедры электронной техники и технологии Белорусского государственного университета информатики и радиоэлектроники. Область научного интереса – компьютерная инженерия, программирование.

А.В. Шкрабов

Студент кафедры электронной техники и технологии Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем.

С.К. Дик

Заведующий кафедрой инженерной и компьютерной графики Белорусский государственный университет информатики и радиоэлектроники, кандидат физико-математических наук, доцент. Окончил Минский радиотехнический институт по специальности «Радиотехника», руководит научными исследованиями в области лазерной медицины.

Г.Д. Ситник

Кандидат медицинских наук, доцент, врач высшей категории по неврологии. Область научного интереса – лечение больных с неврологическими проявлениями поясничного остеохондроза.

И.И. Ревинская

Окончила Белорусский государственный университет информатики и радиоэлектроники. Старший преподаватель кафедры электронной техники и технологии. Область научного интереса – медицинская электроника и обработка медицинских сигналов.

Аннотация. В работе рассматривается подход к повышению устойчивости (robustness) нейросетевых моделей сегментации глаз к типичным искажениям изображений с помощью целенаправленной аугментации данных. На примере победителя OpenEDS Semantic Segmentation Challenge 2019 – модели RitNet (на базе U-Net и DenseNet), показано, как добавление синтетических артефактов (структурированные блики starburst, гауссово размытие, случайные линии и сдвиги), характерных для изображений в VR/AR-очках, позволило достичь высокой точности при крайне малом размере модели. Такой подход подчёркивает важность адаптации обучающих данных под реальные условия применения для повышения надёжности моделей в задачах компьютерного зрения.

Ключевые слова: большие данные, свёрточные нейронные сети, сегментация, устойчивость моделей, инфракрасные камеры, аугментация данных, системы детекции глаз

Введение. В настоящее время алгоритмы на основе нейронных сетей встречаются всё большее применение в различных сферах жизни, в том числе и в среде распознавания образов на изображениях и видео. К лидирующим сферам применения относится здравоохранение, промышленность (выявление дефектов на производстве, анализ инфраструктуры, охрана труда), агрокультура, транспортная сфера, а также видеонаблюдение [2]. Для обучения моделей требуются огромные объёмы данных, однако, существуют некоторые проблемы, связанные с возможной узконаправленностью итогового алгоритма. Модели, обученные на слишком «идеальных» данных, могут резко терять точность при столкновении с неизвестными аномалиями в получаемой информации, а в случае гомогенности данных (например, обучение модели для распознавания лиц на фотографиях людей схожего телосложения, расы или при однотипном освещении) алгоритмы выходят узкоспециализированными, и также перестают правильно обрабатывать информацию при выходе её за известные параметры.

Для борьбы с этим применяются различные методики, обучающие модели разным вариациям данных и помогающие им лучше обрабатывать расхождения в результатах [3-4]. К методам повышения точности машинного обучения можно отнести следующие:

- Модификация (аугментация) данных – преобразование существующих изображений с целью обучения модели на более разнообразных примерах. Используются методики искажения исходных данных, смешивания данных между собой для выделения границ между классами, а также стилизация данных с целью вариативности исходного материала;

- Синтетические данные – искусственное расширение входного массива за счёт данных, сгенерированных иными моделями, что позволяет закрыть пробелы в исходном наборе и внедрить редкие, но возможные сценарии. Тем не менее, этот метод требует тщательного контроля, позволяющего избежать перенесения так называемых «галлюцинаций» между моделями – ложной информации, создаваемых нейросетевыми моделями при недостаточных исходных данных [5];

- Использование дополнительных внешних данных – намеренное добавление изображений иных классов, которые позволят модели чётче определять границы между искомым классом и сторонними данными;

- Обучение работе с неизвестными классами с целью научить модель как узнавать конкретные образы, так и распознавать то, что предоставленный объект не относится ни к одному известному классу;

- Добавление к модели набора признаков других, предварительно обученных моделей с целью расширения набора параметров;

- Активное обучение – ручная коррекция набора данных моделей с помощью ручной разметки пограничных результатов, что может помочь с разрешением неопределённостей;
- Ансамблирование – параллельное использование нескольких моделей с целью агрегации их результатов и коррекции найденных расхождений.

В данной работе сконцентрируемся на первом методе на примере свёрточной нейронной сети RitNet и том, как применение аугментации данных позволило повысить точность и уменьшить размер модели. Полученные результаты иллюстрируют потенциал аугментации данных в задачах Big Data и промышленной аналитики: вместо увеличения объёма «чистых» размеченных данных можно искусственно расширить вариативность существующих наборов, закрывая пробел между лабораторными и реальными условиями. Это особенно актуально для сенсорных данных из IoT, автономного транспорта, здравоохранения и AR/VR, где искажения (размытие, отражения, переменное освещение) являются нормой.

Особенности подхода RitNet. В качестве примера рассмотрим свёрточную модель RitNet [1], разработанную в Рочестерском Институте Технологий (RIT) в 2019 году для семантической сегментации глаза в системах отслеживания взгляда. Модель использует комбинацию архитектуры DenseNet с элементами U-Net и специально взвешенной функции потерь, учитывающей границы. Для обучения модели применялся набор данных OpenEDS [6], упомянутый ранее. Набор состоит из 12759 монохромных изображений глаз людей, полученных с инфракрасных камер, установленных в шлемах виртуальной реальности. Данные получены от 152 субъектов, примерно равно разделённых на мужчин и женщин возрастной группы от 18 до 65 лет. К изображениям прилагается размеченная маска, на которой отмечены зрачок, радужная и белковая оболочка глаз. В конкурсе, для которого была разработана модель, данные были разбиты на три поднабора: обучение (8916 изображений), подтверждение (2403 изображений) и тест (1440 изображений без размеченных масок). Предоставленный набор данных является сравнительно небольшим, и сфокусирован на реальных данных с VR/AR очков и обладает характерными для подобных сценариев артефактами (отражения от подсветки, блики, размытие от движения глаз, частичное перекрытие частей глаза веком). Для оценки качества модели в конкурсе использовалась формула:

$$mIoU + \min(1/S, 1)^2,$$

где $mIoU$ – mean Intersection over Union, или среднее значение перекрытия найденной искомой области реальной (рисунок 1);

S – размер модели (функция от количества параметров модели, в мегабайтах).

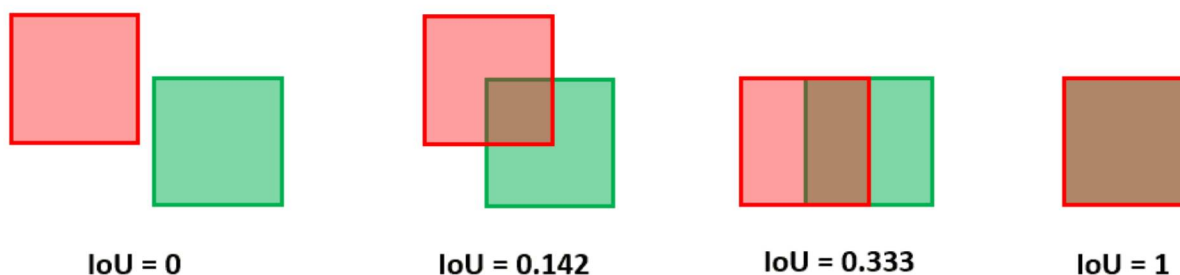


Рисунок 1. Визуальное отображение метрики IoU

Обучение производилось с использованием алгоритма Adam [7] партиями по восемь изображений в сто семьдесят пять проходов (epoch), с модификациями в степени обучения, когда результаты выходили на плато длиной более чем в пять проходов.

Для того, чтобы учесть различия в отражающей способности между образцами в зависимости от цвета глаз, оттенка кожи, а также макияжа и век, изображения, поступающие

в модель для обучения, проходили две ступени предобработки. Первым шагом была экспоненциальная гамма-коррекция изображений, а затем к данным была применено ограниченное контрастом адаптивное выравнивание гистограммы (CLAHE) (рисунок 2).

На преобразованных таким образом изображениях стало проще отделить зрачок от радужки глаза.

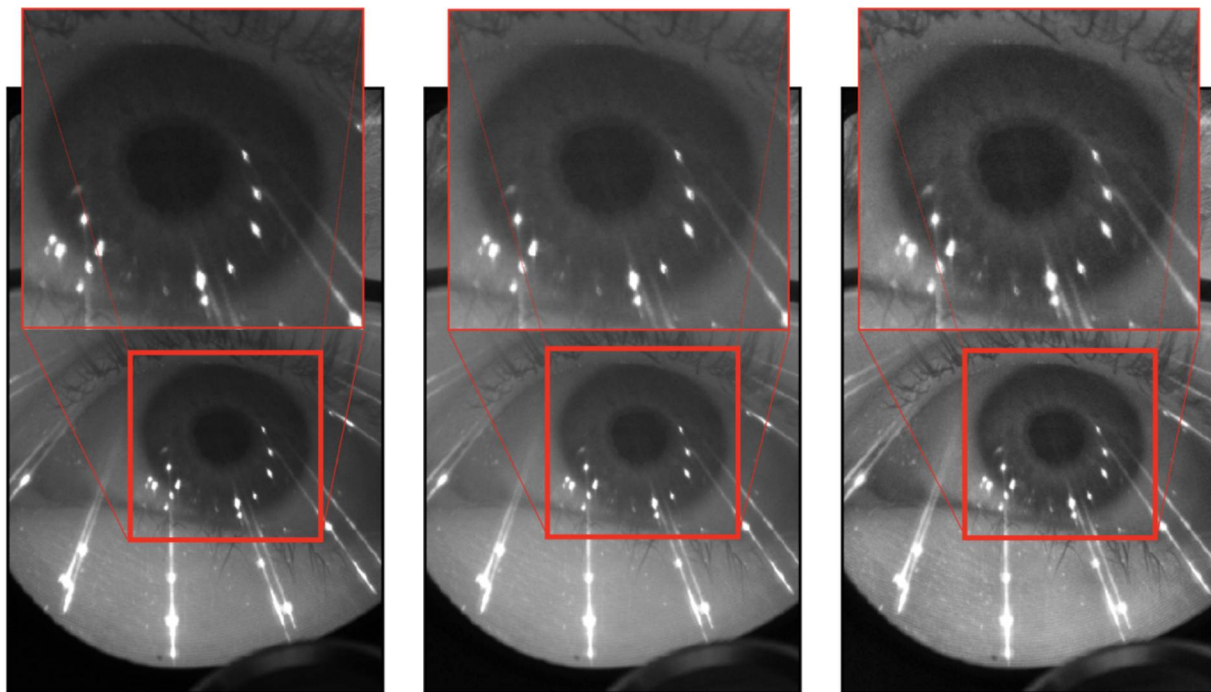


Рисунок 2. Слева направо: оригинальное изображение, изображение после гамма-коррекции, изображение после применения CLAHE

Кроме того, с целью повышения устойчивости модели изображения были подвергнуты следующим аугментациям:

- Отражение по вертикали;
- Размытие по Гауссу с размером ядра 7×7 и случайным отклонением $2 \leq \sigma \leq 7$;
- Сдвиг на 0-20 пикселей по обоим осям;
- Добавление 2-9 случайных тонких линий, построенных от случайного центра ($120 < x < 280$, $192 < y < 448$) для имитации грязи и ресниц;
- Наложение узора (рисунок 3) для уменьшения ошибок сегментации в связи с отражением инфракрасной подсветки от очков. При наложении узор смещался на случайное расстояние от 0 до 40 пикселей по обоим осям.

Каждое изображение было дополнено хотя бы одной из этих аугментаций, при этом появление каждой имело шанс 0.2. Кроме того, каждое изображение было отражено по горизонтали с шансом 0.5 (имитация правого и левого глаза).

Таким образом, в исходные данные были заранее внесены характерные для инфракрасной съемки глаз дефекты, что позволило итоговой модели лучше справляться с ними в дальнейшем.

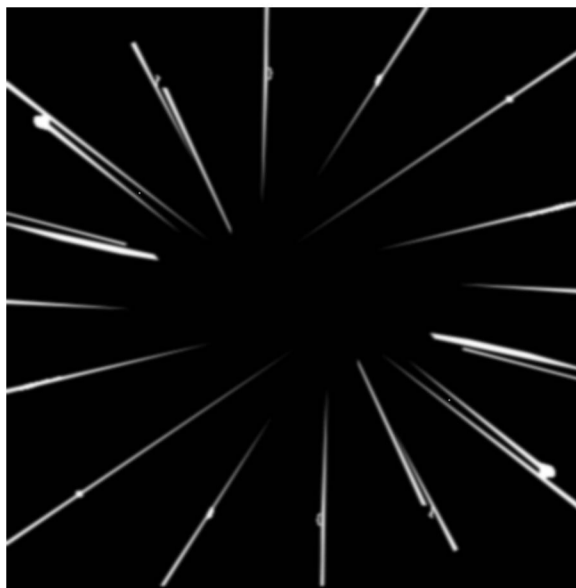


Рисунок 3. Узор, применяемый при аугментации для имитации отражений

Результаты работы. Полученная модель была сравнена с другой свёрточной моделью SegNet [8] (на примере её модификации mSegNet), а также моделями на её основе. К ним относятся:

- mSegNet w/BR (mSegNet с дополнительным модулем уточнения границ (Boundary Refinement, BR), моделирующий выравнивание границ как остаточную (residual) структуру). Такой подход, связанный с принципами остаточного обучения [9], позволяет сети целенаправленно улучшать локализацию границ между классами (в особенности между зрачком, радужкой и склерой), минимизируя ошибки на переходных зонах и повышая общую точность сегментации в условиях классового дисбаланса и низкого контраста.

- mSegNet w/SC – облегчённая версия базовой mSegNet, в которой стандартные свёртки заменены на depthwise separable convolutions (глубинно-разделимые свёртки). Данная техника, существенно снижает количество параметров и вычислительную сложность (примерно в 8–9 раз), сохраняя при этом сопоставимую работоспособность модели.

Сравнительная таблица результатов работы моделей предоставлена в таблице 1. Визуальное отображение работы модели показано на рисунке 4.

Таблица 1. Сравнение параметров различных моделей, участвующих в опыте

Модель	Средний IoU	Размер модели (S), мегабайт	Количество параметров, млн	Итоговый счёт
mSegNet	90,7	13.3	3.5	0.491
mSegNet w/BR	91.4	13.3	3.5	0.495
mSegNet w/SC	89.5	1.6	0.4	0.762
RitNet	95.3	0.98	0.25	0.976

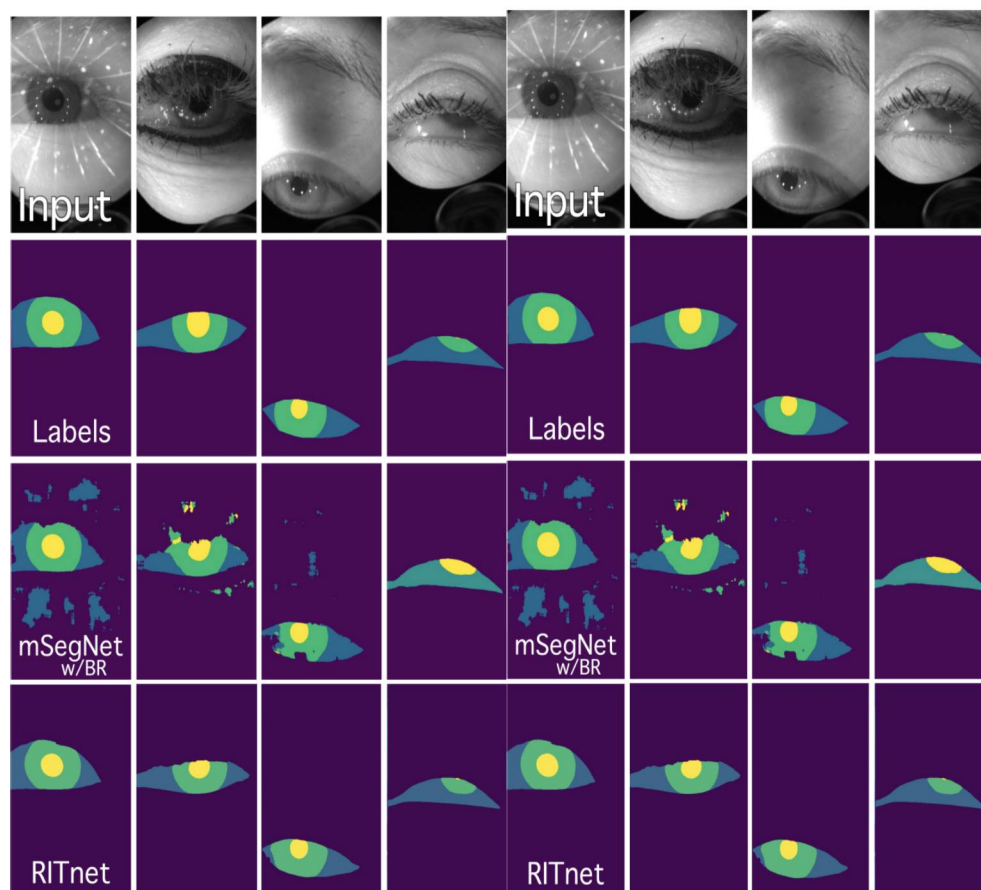


Рисунок 4. Сравнение сегментации между размеченными вручную масками и результатами работы различных моделей. Сверху вниз, первый ряд: изначальные кадры, размеченные вручную маски, второй ряд: результаты работы mSegNet w/BR, результаты работы модели RitNet.

RitNet показывает лучшие значения точности при меньшем размере, чем остальные предоставленные модели. Кроме того, точность классификации тоже выше, чем у аналогов. Небольшой размер модели позволяет использовать её для видео частотой до трёхсот кадров в секунду в реальном времени.

Заключение. Опыт RitNet является наглядным доказательством того, что при работе с обучением моделей качество данных является не менее значимым, чем количество, и, кроме того, показывает, что аугментация является крайне полезным методом повышения устойчивости нейросетевых моделей. С момента релиза RitNet были проведены дальнейшие работы в направлении создания легких и высокоскоростных моделей. Например, в статье Б. Бурлаку [10] сообщается о модели *Lightweight* на основе *MobileNetV2*, обученной на *OpenEDS* с аналогичными аугментациями и показавшей результат сравнимый с RitNet ($mIoU - 0.9485$, конечный счёт – 0.974) с меньшим количеством параметров (104728 против 248900 у RitNet), что обеспечено лучшей архитектурой. Это показывает перспективу и направления для дальнейшей работы в этом направлении.

Современные тенденции [2]) подтверждают, что *data-centric* подходы, включая *domain-specific* аугментацию и лёгкие архитектуры (*MobileNetV2*, *EfficientNet*), остаются ключевыми для промышленного внедрения моделей в условиях ограниченных ресурсов и неидеальных входных данных.

Список литературы

- [1] Chaudhary A. K. [и др.] RitNet: Real-time Semantic Segmentation of the Eye for Gaze Tracking // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). – IEEE, 2019. – С. 3698–3702. – DOI: 10.1109/iccvw.2019.00568.
- [2] Vision AI Trends to Watch in 2026 [Электронный ресурс] // Roboflow Blog. – URL: <https://blog.roboflow.com/vision-ai-trends-2026/> (дата обращения: 10.02.2026).
- [3] Lei S., Hua Y., Zhihao S. Revisiting Fine-Tuning: A Survey of Parameter-Efficient Techniques for Large AI Models [Электронный ресурс]. – 2025. – fhal-05008993. – URL: <https://hal.science/hal-05008993> (дата обращения: 12.02.2026).
- [4] Wang K. [и др.] A Survey on Data Synthesis and Augmentation for Large Language Models [Электронный ресурс] // arXiv preprint. – 2024. – arXiv:2410.12896. – URL: <https://arxiv.org/abs/2410.12896> (дата обращения: 12.02.2026).
- [5] Kalai A. T. [и др.] Why Language Models Hallucinate [Электронный ресурс] // arXiv preprint. – 2025. – arXiv:2509.04664. – URL: <https://arxiv.org/abs/2509.04664> (дата обращения: 12.02.2026).
- [6] Garbin S. J. [и др.] OpenEDS: Open Eye Dataset [Электронный ресурс] // arXiv preprint. – 2019. – arXiv:1905.03702. – URL: <https://arxiv.org/abs/1905.03702>.
- [7] Kingma D. P., Ba J. Adam: A method for stochastic optimization // Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). – San Diego, 2015.
- [8] Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017.
- [9] He K. [и др.] Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – С. 770–778. – DOI: 10.1109/CVPR.2016.90.
- [10] Burlacu B. A. Eye Semantic Segmentation with Convolutional Neural Networks // CS230: Deep Learning, Stanford University, Autumn 2019 [Электронный ресурс]. – URL: https://cs230.stanford.edu/projects_fall_2019/reports/26259743.pdf.

Авторский вклад

Авторы внесли равный вклад в написание статьи

A METHOD FOR INCREASING THE ROBUSTNESS OF A CONVOLUTIONAL NEURAL NETWORK USING DATA AUGMENTATION USING THE EXAMPLE OF AN EYE IMAGE SEGMENTATION MODEL

N.A. Reznikov

*student of the Department of
Electronic Technique and
Technology of BSUIR*

G.D. Sitnik

*PhD, Associate Professor at the Department of
Belarusian State Medical University,
Doctor of the highest category in neurology*

A.V. Shkrabov

*student of the Department of
Electronic Technique and
Technology of BSUIR*

S. K. Dzik

*Chair of the Department of
Engineering and Computer
Graphics, PhD, Associate Professor*

I.I. Revinskaya

*Senior Lecturer at the Department
of Electronic Technique and Technology*

Abstract. The paper discusses an approach to increasing the robustness of neural network models used in eye footage segmentation to typical image distortions using targeted data augmentation. Using the example of the winner of the OpenEDS Semantic Segmentation Challenge 2019 – the RitNet model (based on U-Net and DenseNet), it is shown how the addition of synthetic artifacts (structured starbursts, Gaussian blur, random lines and shifts), characteristic of images in VR / AR glasses, made it possible to achieve high accuracy with an extremely small model size. This approach emphasizes the importance of adapting training data to real-world application conditions to improve the reliability of models in computer vision tasks.

Keywords: big data, convolutional neural networks, segmentation, model robustness, infrared cameras, data augmentation, eye detection systems