

ТЕХНИЧЕСКИЕ НАУКИ



УДК 004.418

Проблемы распознавания символов в автоматизированных системах: анализ ключевых затруднений

П.А. Алексян, А.А. Скляренко

Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация

Аннотация

В статье рассматриваются актуальные вопросы распознавания символьных данных в рамках функционирования автоматизированных систем на примере процесса идентификации государственных регистрационных знаков различных транспортных средств. Автором осуществлен комплексный сравнительный анализ показателей устойчивости классических инструментов оптического распознавания — в частности библиотеки Pytesseract — и прогрессивного нейросетевого подхода, основанного на программной архитектуре YOLOv8. Представлено описание оригинальной методики посимвольной детекции, дополненной последующей программной сортировкой объектов вдоль координатной оси, что обеспечивает сохранение четкой структуры текстовой строки при возникновении значительных геометрических искажений. В рамках экспериментального исследования детально проанализировано влияние ключевых этапов предварительной обработки изображений (разрешение DPI, дескьюинг, процедуры морфологической фильтрации и бинаризации) на итоговую точность и общую скорость функционирования программных алгоритмов. Установлено, что специализированная модель YOLOv8 демонстрирует преимущество в два раза по критерию точности при обработке зашумленных массивов информации и одновременно сохраняет высокие параметры производительности. Выявлен специфический эффект дестабилизации функционирования нейросетевых архитектур в условиях применения избыточной бинаризации входного потока данных.

Ключевые слова: распознавание символов, компьютерное зрение, OCR, YOLOv8, Pytesseract, автоматизированные системы, посимвольная детекция, предобработка изображений

Для цитирования. Алексян П.А., Скляренко А.А. Проблемы распознавания символов в автоматизированных системах: анализ ключевых затруднений. *Молодой исследователь Дона*. 2026;11(5):16–22.

The Problem of Character Recognition in Automated Systems: Analysis of Main Difficulties

Petr A. Alekryan, Anna A. Sklyarenko

Don State Technical University, Rostov-on-Don, Russian Federation

Abstract

The article studies the topical issues of character recognition in automated systems on the example of identifying license plate numbers of various vehicles. The authors have conducted a comprehensive comparative analysis of reliability parameters provided by the traditional optical character recognition engines, notably Pytesseract (Python wrapper), and those provided by the advanced neural network technologies based on the YOLOv8 software architecture. The original character-by-character detection technique with subsequent sorting of objects along an axis of coordinates done by a software has been described. The technique ensures the preservation of a distinct text string structure in case of significant geometric distortions. In the frame of the experimental study, the impact of key image preprocessing stages (DPI resolution, deskewing, morphological filtering, and binarization) on the final accuracy and overall performance of software algorithms has been analysed in detail. It has been found that the specialized YOLOv8 model demonstrates twice better accuracy in processing noisy datasets, as well as maintains high performance rate. A specific destabilizing effect in the work of the neural-network-based architectures, which occurs upon excessive binarization of the input data stream, has been identified.

Keywords: character recognition, computer vision, OCR, YOLOv8, Pytesseract, automated systems, character-by-character detection, image preprocessing

For Citation. Alekryan PA, Sklyarenko AA. The Problem of Character Recognition in Automated Systems: Analysis of Main Difficulties. *Young Researcher of Don*. 2026;11(5):16–22.

Введение. В условиях глобальной цифровизации производственных и транспортных процессов распознавание символов выступает одним из фундаментальных направлений в области компьютерного зрения. Автоматизированные системы идентификации, принимающие решения на основе анализа визуальной информации, находят широкое применение в логистических цепочках, интеллектуальных транспортных сетях, государственном секторе и комплексах безопасности. Эффективность функционирования таких систем напрямую определяется достоверностью данных, извлекаемых на этапе первичной обработки изображений.

Любая ошибка, возникающая на этапе детекции или сегментации символов, неизбежно приводит к формированию «каскадных сбоев» в последующих модулях логической обработки. В контексте идентификации государственных регистрационных знаков транспортных средств некорректное считывание даже одного символа делает невозможной автоматическую проверку по базам данных, что требует участия оператора и снижает преимущества автоматизации. В связи с этим минимизация рисков и поиск наиболее устойчивых стратегий распознавания в условиях реальной эксплуатации представляют собой актуальную научно-техническую задачу.

Основная проблема современных алгоритмов распознавания заключается в их уязвимости перед дестабилизирующими факторами окружающей среды. В отличие от лабораторных условий или обработки сканированных документов, системы реального времени сталкиваются с рядом затруднений:

- искажением перспективы и значительным наклоном объекта относительно оптической оси камеры;
- избыточной освещенностью, включая засветы, либо, напротив, низкой контрастностью изображения;
- визуальным перекрытием элементов и наличием специфических шумов.

Традиционные методы оптического распознавания символов (далее — OCR), ярким представителем которых является библиотека Pytesseract, основанная на движке Tesseract, на протяжении длительного времени рассматривались как индустриальный стандарт [1]. Однако жесткая привязка данных методов к структуре строки и их высокая чувствительность к качеству бинаризации изображения ставят под сомнение эффективность применения в сложных эксплуатационных условиях. В качестве альтернативного решения рассматриваются современные нейросетевые архитектуры, в частности детекторные модели семейства YOLO (You Only Look Once) [2].

В данной работе предлагается подход, основанный на использовании архитектуры YOLOv8, обученной для посимвольной детекции. Особенность предлагаемого метода состоит в отказе от распознавания всей строки целиком в пользу идентификации отдельных графических образов символов с их последующим программным ранжированием по координатной оси для формирования итоговой последовательности. Предполагается, что данный подход позволит нивелировать проблему потери структуры данных при геометрических искажениях кадра.

Цель исследования заключается в проведении сравнительного анализа устойчивости классических методов OCR на примере библиотеки Pytesseract и нейросетевого подхода на базе модели YOLOv8 при решении задачи идентификации символьной информации в условиях типовых визуальных помех.

Для достижения поставленной цели необходимо решить следующие задачи:

- оценить точность детекции символов обоими методами на исходных изображениях без предварительной обработки;
- проанализировать влияние различных этапов программной предобработки, включая изменение DPI, выравнивание горизонта и морфологическую фильтрацию, на результирующие показатели точности и скорости работы алгоритмов;
- выявить ключевые факторы, вызывающие ложные срабатывания и отказы в работе исследуемых систем.

Основная часть. Ключевым отличием предложенной методики применения YOLOv8 является отказ от стандартной парадигмы распознавания всей текстовой строки целиком. В рамках настоящего исследования реализован подход посимвольной детекции, при котором модель обучалась идентификации отдельных графических образов символов — букв и цифр — как независимых объектов в кадре. После этапа нейросетевой детекции для получения строки автомобильного номера используется алгоритм постобработки, который выполняет извлечение координат ограничивающих рамок (bounding boxes) для каждого обнаруженного символа, сортировку объектов по горизонтальной координатной оси X и последующую агрегацию предсказанных меток классов в единую строку.

Пример логики работы данного метода представлен на рис. 1.

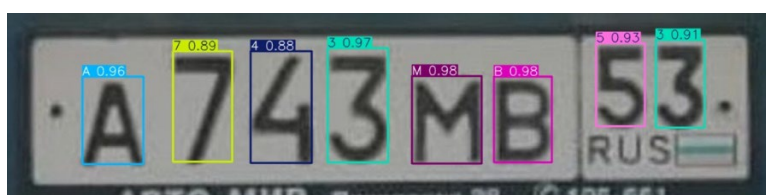


Рис. 1. Логика работы посимвольной детекции и последующей сортировки

Так как чтение строки целиком в классическом OCR-решении полностью зависит от исходной структуры текстовой последовательности на изображении, основное преимущество метода посимвольной детекции проявляется в его устойчивости к нарушению этой структуры. Даже при наличии геометрических перекосов, частичном перекрытии сторонними объектами или иных помехах на изображении автомобильного номера данный подход опирается на точные координаты каждого символа в отдельности. После последующего восстановления последовательности по координате X формируется итоговая строка.

Для проведения эксперимента был собран датасет, включающий 500 изображений государственных регистрационных знаков. Каждое изображение прошло ручную разметку с выделением отдельных независимых классов для каждого символа. При обучении модели не применялись методы аугментации, в том числе геометрические преобразования (масштабирование, сдвиги, обрезка и т. д.) и фотометрические преобразования (коррекция яркости, контрастности и т. д.), поскольку исследование предполагает оценку адаптивности архитектуры YOLOv8 в условиях работы с ограниченными выборками данных.

Для оценки влияния визуальных искажений на оба метода была сформирована тестовая выборка с учетом основных условий, воздействующих на точность детекции строки. Использование единой тестовой выборки позволяет отдельно оценить вклад каждого типа визуального искажения в итоговые метрики точности распознавания. Перечень таких искажений представлен в таблице 1.

Таблица 1

Перечень визуальных искажений

Тип	Описание	Влияние
Нарушение геометрии изображения	Поворот номерного знака относительно горизонтальной оси	Снижение точности распознавания строки целиком
Световые искажения	Чрезмерное освещение или наличие засветов	Снижение контрастности фона и символов
Внешние шумы	Условия плохой видимости или загрязнение	Снижение чёткости контуров символов
Частичное перекрытие	Посторонние предметы в объективе (прикрытый номер, проходящий мимо человек)	Потеря символа/ов из изображения

Пример изображений с подобными визуальными искажениями отображен на рис. 2.

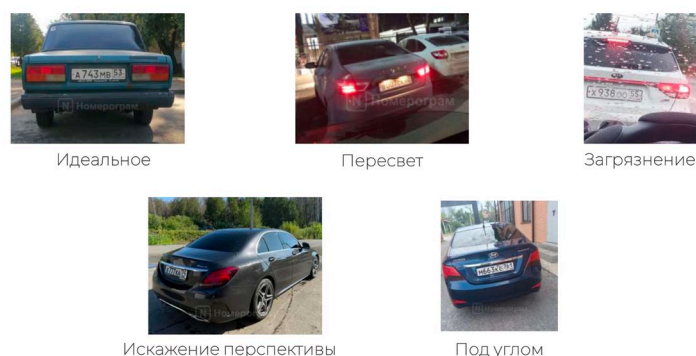


Рис. 2. Примеры тестовых изображений с различными типами помех

Для оценки точности и эффективности исследуемых методов в рамках данной работы были определены три ключевых показателя. Базовым индикатором выступала посимвольная точность, которая отражала отношение количества правильно идентифицированных графических знаков к их общему числу в строке. В приоритете же для систем контроля и управления доступом находился показатель точности распознавания всей строки целиком, фиксирующий долю безошибочно считанных государственных регистрационных знаков относительно общего объема тестовой выборки. Дополнительно производилось измерение среднего времени инференса на один кадр, что позволило проанализировать вычислительную сложность каждого из рассматриваемых подходов.

Исследование базируется на детальном анализе влияния различных этапов программной предобработки входного изображения на итоговые качественные и скоростные показатели методов. Выбранная стратегия позволила наглядно продемонстрировать условия, при которых алгоритмы способны дифференцировать образы символов на зашумленном фоне, а также выявить моменты снижения точности вследствие критических искажений.

Пилотная проверка осуществлялась на исходных изображениях, содержащих естественные засветы, загрязнения, эффекты искаженной перспективы и автомобильные номера, расположенные под значительным углом. Классическое OCR-решение обеспечило корректное считывание отдельных символов и строки в целом лишь в единичных случаях, что подтвердило его уязвимость и низкую эффективность детекции на изображениях без предварительной подготовки. В то же время нейросетевая модель YOLOv8 продемонстрировала способность к успешной идентификации как независимых знаков, так и последовательностей символов даже в сложных условиях. Данный факт подтверждает, что метод на базе архитектуры YOLOv8 обладает существенным преимуществом перед традиционным OCR в ситуациях, характеризующихся деформацией текстовой строки и наличием интенсивных внешних помех. Результаты функционирования алгоритмов на необработанных данных представлены на рис. 3.

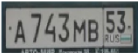
Изображение	Реальное значение	YOLO	Pytesseract
	A743MB53	A743MB53 (126.5 ms)	A743MB23 (305.2 ms)
	B007CK02	B0015K2 (40%) (70.1 ms)	- (190.4 ms)
	X9380055	X938055 (93%) (52.9 ms)	X93859553 (199.0 ms)
	E777KA74	E7771K174 (77%) (70.3 ms)	P (231.8 ms)
	M663KE761	M663KE761 (63.8 ms)	66KE (217.7 ms)

Рис. 3. Сравнение работы алгоритмов на исходных изображениях

Программное увеличение разрешения до 300 DPI было сделано для проверки гипотезы о необходимости повышения плотности пикселей для улучшения сегментации. Результаты показали, что для YOLOv8 данная операция практически не влияет на точность, но сохраняет высокую скорость обработки. Для OCR увеличение объема данных привело к значительному замедлению обработки кадра, что делает такую предобработку сомнительной для систем, работающих в режиме реального времени. Результаты работы алгоритмов на данных с увеличенным разрешением изображены на рис. 4.

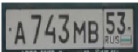
Изображение	Реальное значение	YOLO	Pytesseract
	A743MB53	A743MB53 (135.6 ms)	A743MB23 (336.1 ms)
	B007CK02	B0015K (28%) (63.3 ms)	- (323.7 ms)
	X9380055	X938055 (93%) (60.3 ms)	X93890553 (370.8 ms)
	E777KA74	E7771K1774 (73%) (69.5 ms)	B (363.4 ms)
	M663KE761	M663KE761 (55.0 ms)	- (424.2 ms)

Рис. 4. Влияние изменения DPI на скорость и точность

На этапе выравнивания горизонта устранялся наклон государственного регистрационного знака относительно горизонтальной оси. Визуальный анализ показал, что именно данный этап стал первым, на котором классическое OCR-решение смогло частично восстановить линейную структуру строки и выполнить считывание отдельных символов. Для модели YOLOv8 такая процедура также оказалась значимой, поскольку после коррекции положения номера были устранены артефакты детекции, обусловленные поворотом ограничивающих рамок. При этом сохранялись отдельные ошибки, связанные с визуальным сходством некоторых символов, например, «7» и «У». Результаты функционирования алгоритмов на данных с выровненным горизонтом представлены на рис. 5.

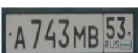
Изображение	Реальное значение	YOLO	Pytesseract
	A743MB53	A743MB53 (120.5 ms)	A743MB23 (300.6 ms)
	B007CK02	B0075K02 (50%) (71.0 ms)	- (333.5 ms)
	X9380055	X93855 (85%) (53.1 ms)	93820553 (348.1 ms)
	E777KA74	ET7YKA74 (75%) (76.8 ms)	- (351.5 ms)
	M663KE761	M663KE761 (58.9 ms)	MOO3KE76 (399.0 ms)

Рис. 5. Результаты распознавания после коррекции угла наклона

Использование морфологических преобразований, в частности эрозии и дилатации, было направлено на повышение выраженности контуров символов и усиление их отделимости от фона. Данная мера оказалась наиболее обоснованной для кадров с выраженной переэкспозицией, где границы знаков становились менее четкими вследствие интенсивных засветов. На этом этапе был зафиксирован устойчивый рост точности OCR, тогда как для YOLOv8 наблюдался незначительный обратный эффект. Морфологическая обработка изменяла характерные контуры и признаки символов, на которых была обучена нейросетевая модель, поэтому она начинала менее уверенно дифференцировать образы и допускала ошибки распознавания. В то же время для OCR уменьшалось количество разрывов в линиях, а контраст между символами и фоном становился более бинарным и однородным, что обеспечивало более точную идентификацию строки данным методом. Результаты функционирования алгоритмов на данных с применением морфологической обработки представлены на рис. 6.

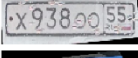
Изображение	Реальное значение	YOLO	Pytesseract
	A743MB53	A743MB53 (114.9 ms)	A743MB23 (318.9 ms)
	B007CK02	B007CK02 (68.4 ms)	B07CK (61%) (280.7 ms)
	X9380055	X933855 (80%) (51.3 ms)	93850553 (338.0 ms)
	E777KA74	E77YKA74 (75%) (75.3 ms)	- (339.1 ms)
	M663KE761	M663KE761 (58.4 ms)	MO663KE761 (322.7 ms)

Рис. 6. Адаптация алгоритмов к морфологическим изменениям кадра

В качестве заключительного этапа изображения были приведены к бинаризованному, то есть черно-белому, формату. Несмотря на то, что данный вариант обработки теоретически должен формировать наиболее благоприятные условия для OCR, алгоритм не продемонстрировал стабильных результатов не только при распознавании сложных регистрационных знаков, но и при обработке кадров без выраженных помех. Модель YOLOv8 также показала снижение метрик, поскольку в ее обучающей выборке отсутствовали бинаризованные изображения. В результате нейросетевая модель утратила доступ к привычным визуальным признакам, таким как текстура, объем и цвет, вследствие чего не смогла эффективно функционировать в новых условиях представления данных. Результаты работы алгоритмов на бинаризованных данных представлены на рис. 7.

Изображение	Реальное значение	YOLO	Pytesseract
	A743MB53	A743MB53 (116.8 ms)	A743MB23 (233.8 ms)
	B007CK02	B0BCK (46%) (70.5 ms)	A (260.3 ms)
	X9380055	X9380055E (94%) (52.0 ms)	93850884 (259.7 ms)
	E777KA74	E7K74 (76%) (73.9 ms)	ETT (261.0 ms)
	M663KE761	M66E3MKE761 (90%) (62.5 ms)	MO63KE701 (251.2 ms)

Рис. 7. Сравнение устойчивости методов в условиях бинаризации

Итоговое сопоставление эффективности исследуемых методов проводилось на контрольной выборке, включавшей 125 изображений — по 5 снимков для каждой категории помех. Выполненный количественный и качественный анализ позволил выявить ряд закономерностей в функционировании систем:

1. Статистический анализ показал существенное преимущество нейросетевого подхода по количественным показателям точности. На зашумленных выборках средняя посимвольная точность YOLOv8 практически вдвое превысила результаты Pytesseract. Если классический OCR-метод сохраняет работоспособность преимущественно в условиях, близких к лабораторным, то детекторная модель демонстрирует высокую устойчивость к внешним искажениям. Наиболее показательной стала метрика распознавания строки целиком. В условиях реальных помех Pytesseract не обеспечил стабильного результата, поскольку даже незначительный сбой сегментации хотя бы одного символа приводил к невалидности всей последовательности. Внедрение предложенного алгоритма программной сортировки координат позволило YOLOv8 корректно восстанавливать структуру номера даже при значительном нарушении визуальной связности строки. Результаты сравнения точности представлены на рис. 8.

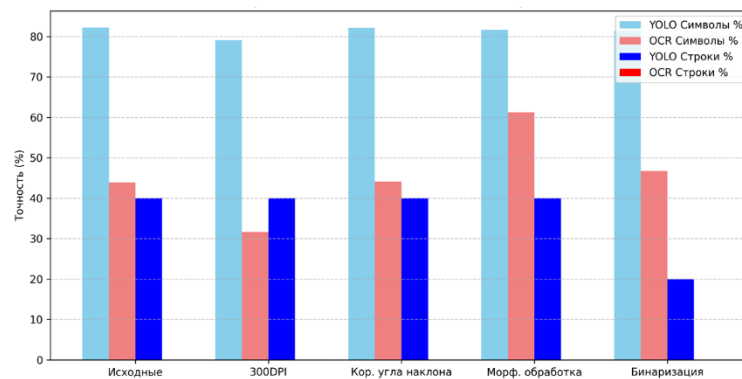


Рис. 8. Сравнение методов по точности распознавания

2. Анализ времени обработки одного кадра показал, что YOLOv8 работает значительно быстрее. Среднее время инференса у нейросети оказалось в разы ниже, чем у классического OCR. При этом средняя скорость YOLOv8 остается стабильной и не сильно зависит от того, какие фильтры предобработки применялись к кадру. С OCR нагрузка напрямую зависит от входных данных. Например, если программно поднять разрешение до 300 DPI, время обработки начинает расти экспоненциально. Такая задержка делает классические методы непригодными для высоконагруженных систем, где нужно анализировать видеопоток в реальном времени. Сравнение скорости работы алгоритмов приведено на рис. 9.

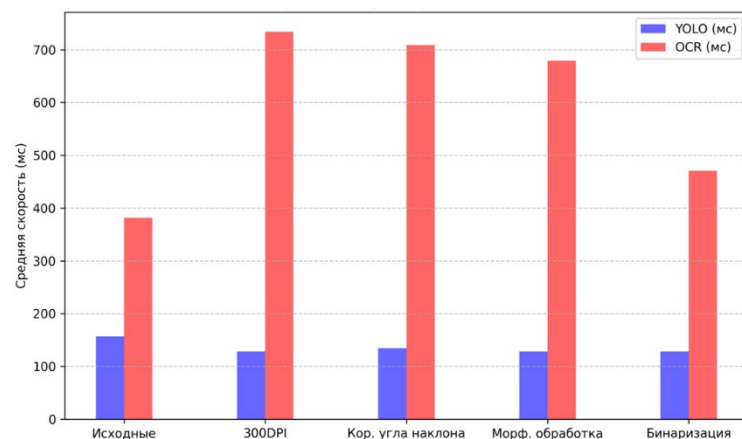


Рис. 9. Сравнение методов по скорости распознавания

Операции предварительной обработки являются обязательными для OCR-решений, поскольку они направлены на приведение входного сигнала к строгим стандартам сегментации, заложенным в логику работы соответствующих алгоритмов. В то же время для глубоких нейронных сетей чрезмерная очистка данных нередко становится дестабилизирующим фактором. Поскольку архитектура YOLOv8 обучалась на репрезентативной выборке реальных изображений, переход к «стерильным» бинаризованным формам лишает модель естественных текстурных признаков и полутонов. Подобная трансформация привычных паттернов объекта приводит к увеличению числа ложных срабатываний и подтверждает, что нейросетевой подход требует сохранения контекстной информации, которую классические системы, как правило, интерпретируют как шум.

Тем не менее оба метода сохраняют чувствительность к визуально сходным символам. Наиболее распространенной ошибкой в рамках эксперимента стала путаница между символом «7» и кириллической буквой «У» в условиях низкой контрастности. Однако, в отличие от Pytesseract, который в подобных ситуациях часто полностью теряет структуру строки, YOLOv8 фиксирует наличие объекта, что позволяет в дальнейшем применять вероятностные фильтры для коррекции результата. Общая структура выводов о точности алгоритмов представлена на рис. 10.



Рис. 10. Анализ устойчивости методов к различным типам помех

Заключение. В ходе проведенного исследования был выполнен сравнительный анализ устойчивости классических и нейросетевых методов распознавания символов в условиях, имитирующих реальную эксплуатацию автоматизированных систем.

На основе полученных данных можно сформулировать следующие выводы:

– Комбинация YOLOv8 и алгоритма сортировки символов по оси X позволила повысить точность распознавания примерно в 2 раза, а также обеспечить абсолютное преимущество при идентификации всей строки: OCR не смог корректно распознать строку целиком ни на одном кадре, тогда как YOLOv8 успешно справилась приблизительно с 42 % таких задач.

– Нейросетевые модели демонстрируют значительно более стабильную работу на кадрах с наклоном или нарушенной перспективой. Классический OCR критически зависит от идеального выравнивания строки, поэтому любой перекос существенно снижает качество распознавания. В то же время YOLOv8 не имеет подобного ограничения, поскольку успешно обнаруживает и распознает каждый символ как отдельный независимый объект.

– Анализ скорости работы подтвердил, что нейросетевые подходы хорошо подходят для систем реального времени. YOLOv8 обрабатывает кадр в несколько раз быстрее, чем PyTesseract. При этом производительность нейросетевой модели снижается лишь незначительно при увеличении разрешения входных изображений.

– Важным научным результатом работы является выявление негативного влияния избыточной фильтрации на точность нейронных сетей. Если для классических алгоритмов бинаризация и морфологическая обработка являются необходимыми этапами, то для современных детекторов они могут выступать в качестве дестабилизирующего фактора, искажая значимые признаки объектов.

Практическая значимость работы заключается в подтверждении возможности создания высокоточных систем идентификации даже на относительно небольших обучающих выборках объемом 500 экземпляров при условии применения современных архитектур компьютерного зрения.

Дальнейшее развитие исследования связано с расширением обучающего набора данных за счет использования методов аугментации: внесения синтетических шумов, имитации различных погодных условий и типов освещения. Это позволит дополнительно повысить устойчивость системы к специфическим визуальным ошибкам и минимизировать риск «каскадных сбоев» в промышленных и транспортных автоматизированных комплексах.

Список литературы

1. *Оптическое распознавание символов для Python (PyTesseract)*. URL: <https://pytesseract.readthedocs.io/en/latest> (дата обращения: 31.03.2026).
2. *Документация нейросетевой архитектуры YOLOv8*. URL: <https://docs.ultralytics.com/models/yolov8> (дата обращения: 31.03.2026).

Об авторах:

Петр Артемович Алекьян, магистрант кафедры «Программное обеспечение вычислительной техники и автоматизированных систем» Донского государственного технического университета (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1), alekolen35@gmail.com

Анна Анатольевна Скляренко, доцент, кандидат технических наук, доцент кафедры «Программное обеспечение вычислительной техники и автоматизированных систем» Донского государственного технического университета (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1) Deranna1982@yandex.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the Authors:

Petr A. Alekya, Master's Degree Student of the Software of Computer Engineering and Automated Systems Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation), alekolen35@gmail.com

Anna A. Sklyarenko, Cand.Sci. (Engineering), Associate Professor of the Software of Computer Engineering and Automated Systems Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation) Deranna1982@yandex.ru

Conflict of Interest Statement: the authors declare no conflict of interest.

All authors have read and approved the final manuscript.