

Подсистема классификации объектов в составе комплекса обнаружения и сопровождения БПЛА на основе методов компьютерного зрения

Д. М. Сошников

Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых
600000, г. Владимир, ул. Горького, 87
E-mail: fimov41@mail.ru

Предложен подход к классификации изображений на основе глубоких сверточных нейронных сетей с последующей интеграцией в прикладное программное обеспечение. В рамках работы проведено сравнительное исследование современных архитектур, включая ResNet-50, ResNeXt-50 и EfficientNet, применяемых для задачи классификации объектов. По результатам экспериментов наилучшую точность продемонстрировала модель ResNeXt-50, достигшая значений до 91,35%, однако она характеризуется повышенными вычислительными затратами. Модель EfficientNet показала высокую эффективность с точки зрения соотношения точности и количества параметров, но оказалась менее стабильной при дообучении на пользовательском датасете ограниченного объема. Наиболее сбалансированные результаты продемонстрировала архитектура ResNet-50, обеспечившая высокую точность классификации, устойчивость к изменениям гиперпараметров и стабильную сходимость в процессе обучения. Реализация подсистемы классификации выполнена как часть общего приложения, разработанного на языке C# с использованием платформы .NET. Обучение модели осуществляется в среде Python с применением библиотеки PyTorch, после чего модель экспортируется в формат ONNX для последующей интеграции. Инференс выполняется непосредственно в C#-приложении с использованием ML.NET, что позволяет обеспечить высокую скорость работы и отказаться от использования Python на этапе эксплуатации. Таким образом, разработанная подсистема классификации обеспечивает эффективное решение задачи распознавания объектов и может быть использована в составе прикладных систем компьютерного зрения, работающих в режиме реального времени.

Ключевые слова: классификация изображений, компьютерное зрение, глубокое обучение, сверточные нейронные сети, ResNet, ResNeXt, EfficientNet, архитектуры нейронных сетей, обработка изображений, распознавание объектов, обучающие выборки, датасеты, инференс, экспорт моделей, интеграция нейросетей, искусственный интеллект, машинное обучение.

Object classification subsystem as part of a UAV detection and tracking system based on computer vision methods

D. M. Soshnikov

Vladimir State University named after Alexander and Nikolai Stoletovs, 87 Gorky Street, Vladimir, 600000, Russia. E-mail: [specify authors' e-mail]

A comprehensive approach to image classification based on deep learning and computer vision is proposed, including model selection, training, evaluation, and integration into application software. A comparative study of modern neural network architectures, including ResNet-50, ResNeXt-50, and EfficientNet, was conducted. ResNeXt-50 achieved the highest classification accuracy of up to 91.35%, while EfficientNet demonstrated a strong balance between accuracy and model complexity but showed lower stability on a limited custom dataset. ResNet-50 provided the most balanced results in terms of accuracy, training stability, and computational efficiency, making it the most suitable choice for practical deployment. The classification model was trained in a Python-based environment and subsequently adapted for integration into a C# application. The trained network was exported to the

ONNX format, enabling inference directly within the .NET environment without requiring Python at runtime. This approach ensures efficient deployment, high performance, and seamless integration into the overall software system.

Keywords: image classification, computer vision, deep learning, convolutional neural networks, ResNet, ResNeXt, EfficientNet, image processing, object recognition, datasets, neural network architectures, artificial intelligence, machine learning.

Введение

В условиях стремительного развития беспилотных летательных аппаратов существенно возрастает необходимость создания интеллектуальных систем, способных не только обнаруживать воздушные цели, но и корректно их классифицировать.

Классификация является ключевым этапом обработки визуальной информации, поскольку именно она позволяет определить принадлежность объекта к конкретному классу и, тем самым, обеспечить основу для принятия решений в системах мониторинга и противодействия. В практических сценариях это может означать различие между потенциально опасным БПЛА, гражданским дроном или посторонним объектом, не представляющим угрозы.

Особенность задачи классификации в рассматриваемой предметной области заключается в высокой неопределённости входных данных. Объекты могут иметь малые размеры относительно кадра, появляться на сложном фоне городской или природной среды, частично перекрываться другими объектами, а также наблюдаться при различных погодных условиях и уровнях освещённости. Дополнительную сложность вносит разнообразие самих БПЛА: различия в конструкции, форме, цвете и динамике движения требуют от модели высокой обобщающей способности.

Традиционные алгоритмы обработки изображений, основанные на ручном выделении признаков, в таких условиях оказываются недостаточно гибкими и плохо масштабируются при увеличении числа классов и вариативности данных. В связи с этим в последние годы широкое распространение получили методы глубокого обучения, в частности сверточные нейронные сети, которые способны автоматически извлекать информативные признаки непосредственно из изображений. Их применение позволяет значительно повысить точность классификации и устойчивость к шумам и искажениям.

Таким образом, актуальной задачей является выбор архитектуры модели классификации, обеспечивающей оптимальный баланс между точностью, устойчивостью к внешним условиям и вычислительной эффективностью. В рамках данной работы особое внимание уделяется архитектурам семейства ResNet, а также их современным модификациям — ResNeXt и EfficientNet, которые представляют различные подходы к построению и масштабированию глубоких нейронных сетей.

В заключение проводится сравнительный анализ указанных архитектур с точки зрения ключевых характеристик: точности классификации, устойчивости к вариативности входных данных, вычислительных затрат и пригодности для работы в реальном времени. Такой анализ позволяет обоснованно выбрать наиболее подходящую модель для интеграции в разрабатываемую систему и определить возможные направления её дальнейшей оптимизации.

Обзор существующих решений

Современные подходы к классификации изображений в задачах компьютерного зрения базируются на использовании глубоких сверточных нейронных сетей, способных автоматически извлекать информативные признаки из входных данных. Это особенно важно в задачах распознавания БПЛА, где объекты могут существенно различаться по форме, размеру и условиям наблюдения. Среди множества архитектур особое

распространение получили модели семейства ResNet, а также их последующие модификации — ResNeXt и EfficientNet.

Архитектура ResNet (Residual Network) представляет собой усовершенствованную модель сверточной нейронной сети, предназначенную для обучения глубоких структур без деградации качества. Её ключевая особенность заключается в использовании остаточных связей (skip connections), позволяющих передавать входной сигнал напрямую через один или несколько слоёв. Базовым элементом сети является остаточный блок, в котором результат последовательных свёрточных преобразований складывается с исходным входом перед применением функции активации. Такой подход позволяет сети «игнорировать» слои, не вносящие вклад в улучшение результата, и существенно упрощает процесс обучения.

Использование остаточных блоков даёт ряд важных преимуществ. Во-первых, значительно снижается проблема затухания градиента, что позволяет эффективно обучать сети глубиной в десятки и сотни слоёв. Во-вторых, обеспечивается более стабильная и быстрая сходимость в процессе обучения, так как градиент может напрямую распространяться к ранним слоям сети. В-третьих, повышается общая устойчивость модели к переобучению и искажениям входных данных.

Дополнительным преимуществом ResNet является способность формировать глубокую иерархию признаков. На начальных слоях извлекаются простые характеристики изображений, такие как границы, текстуры и углы, тогда как более глубокие слои объединяют их в сложные структуры и объекты. Это позволяет эффективно различать даже визуально схожие классы, что критически важно для задач классификации БПЛА. При этом архитектура остаётся достаточно гибкой: существуют различные варианты (ResNet-50, ResNet-101, ResNet-152), позволяющие выбирать баланс между точностью и вычислительными затратами в зависимости от требований задачи.

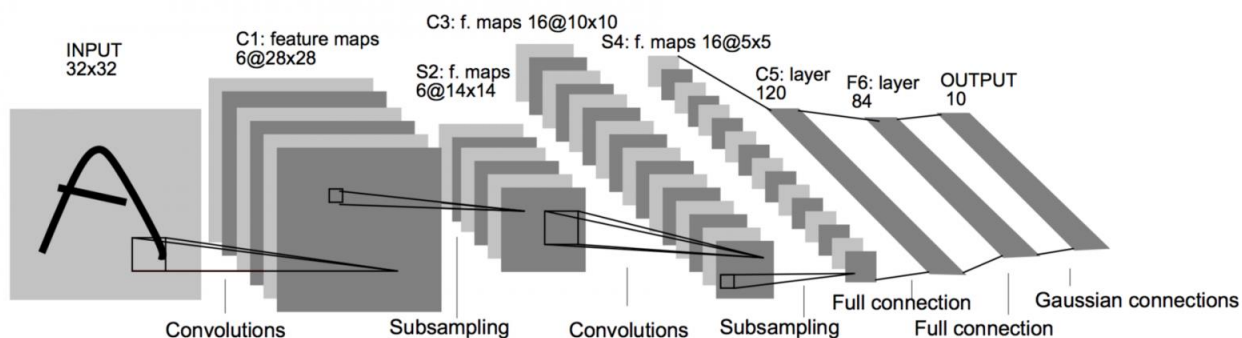


Рис. 1. «Принцип работы ResNet»

Развитием идей ResNet является архитектура ResNeXt, в которой используется принцип агрегированных преобразований. В отличие от классического увеличения глубины или ширины сети, здесь вводится параметр кардинальности — количество параллельных путей внутри одного блока. Это позволяет повысить выразительную способность модели и улучшить точность классификации без существенного роста числа параметров. Однако такая структура усложняет настройку модели и увеличивает требования к вычислительным ресурсам.

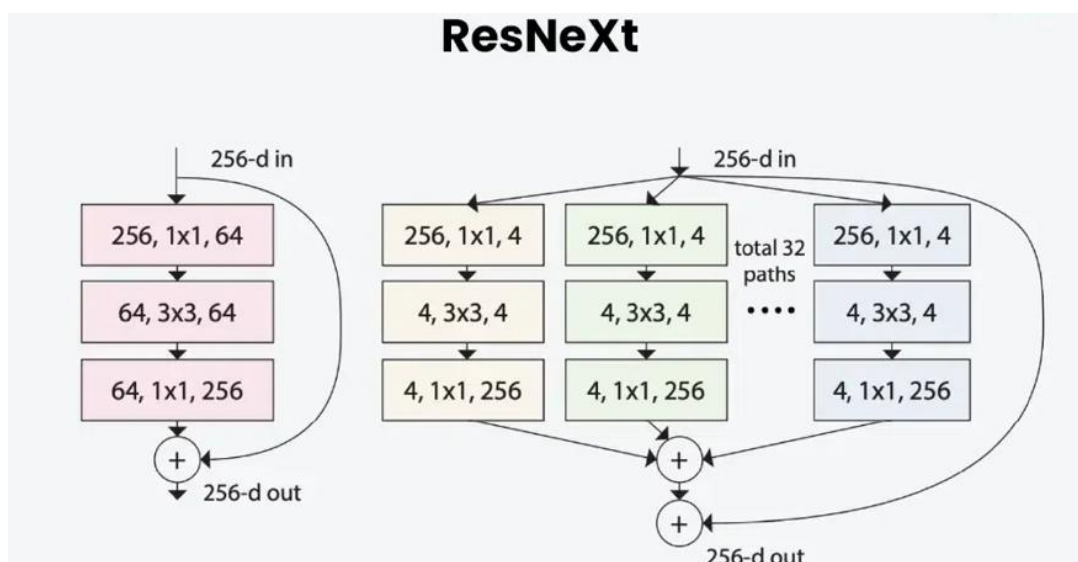


Рис. 2. «Принцип работы ResNeXt»

Альтернативным подходом является семейство моделей EfficientNet, основанное на принципе сбалансированного масштабирования сети. В данной архитектуре одновременно масштабируются глубина, ширина и разрешение входных изображений по специально подобранным коэффициентам. Это позволяет достигать высокой точности при меньшем количестве параметров и вычислений. EfficientNet демонстрирует высокую эффективность, однако в ряде случаев, особенно при дообучении на относительно небольших пользовательских датасетах, её поведение может быть менее стабильным.

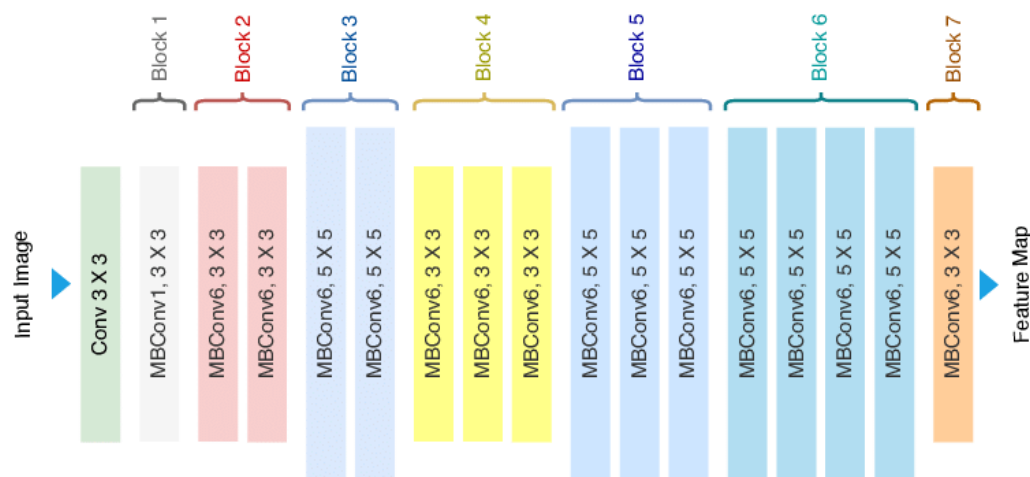


Рис. 3. «Принцип работы EfficientNet»

В рамках проведённого сравнительного анализа указанных архитектур оценивались такие характеристики, как точность классификации, скорость сходимости, устойчивость к переобучению и удобство интеграции в существующий программный пайплайн. Все модели обучались в сопоставимых условиях, что позволило получить объективные результаты. Архитектура ResNeXt показала высокую точность, но оказалась более ресурсоёмкой и сложной в настройке. EfficientNet продемонстрировала хорошее соотношение точности и количества параметров, однако уступила по стабильности при работе с ограниченными данными.

Наиболее сбалансированные результаты продемонстрировала ResNet. Данная архитектура обеспечила стабильное обучение, предсказуемое поведение при изменении гиперпараметров и высокую точность классификации. Дополнительными преимуществами стали простота модификации (в частности, адаптация выходного слоя под число классов), хорошая документированность и удобство интеграции в программную систему. Важным фактором также стала возможность надёжного экспорта модели в формат ONNX и её последующего использования в C#-приложении без потери производительности.

Таким образом, несмотря на наличие более современных архитектур, именно ResNet является наиболее целесообразным выбором для рассматриваемой задачи, поскольку она обеспечивает оптимальное сочетание точности, стабильности и удобства практического применения.

Методология исследования

Задача классификации типов БПЛА рассматривалась как отдельный этап после детекции. Такой подход позволяет не перегружать детектор многоклассовой задачей и обеспечить более высокую точность распознавания типов целей. Для решения этой задачи были выбраны архитектуры ResNet, ResNeXt и EfficientNet, хорошо зарекомендовавшие себя в задачах чистой классификации изображений.

ResNet использует остаточные связи, что облегчает обучение глубоких сетей и обеспечивает устойчивую сходимость. ResNeXt развивает идеи ResNet за счет увеличения кардинальности блоков и позволяет повысить качество при умеренном росте числа параметров. EfficientNet использует согласованное масштабирование глубины, ширины и разрешения входного изображения, обеспечивая высокую эффективность с точки зрения соотношения качества и вычислительной сложности. В обучающем датасете использовались классы delta-wing, fixed-wing, hexacopter, quadrocopter и not-drone. Дополнительно для стабилизации предсказаний применялась пакетная обработка кадров с усреднением вероятностей.

Таблица 1. Сравнение алгоритмов трекинга OpenCV

Трекер	Скорость (FPS)	Точность	Устойчивость к перекрытиям	Сложность настройки
CSRT	20–30	Высокая	Высокая	Низкая
KCF	50–70	Средняя	Средняя	Низкая
MOSSE	100–150	Низкая	Низкая	Низкая
MIL	20–40	Средняя	Средняя	Средняя
GOTURN	20–30	Средняя	Низкая	Высокая

Таблица 2. Сравнение нейросетей

Архитектура	Основная идея	Преимущества	Недостатки
ResNet (Residual Network)	Использует остаточные блоки (skip connections) для предотвращения затухания градиента и эффективного обучения глубоких сетей.	- Высокая точность на больших и средних датасетах.- Стабильное и быстрое обучение глубоких моделей.- Устойчивость к искажениям и сдвигам объектов.- Простота архитектуры и широкая поддержка библиотек.	- Большое количество параметров при очень глубокой сети.- Могут требоваться большие вычислительные ресурсы для

			ResNet-101 и выше.
ResNeXt	Развивает идею ResNet, добавляя кардинальность — несколько параллельных путей внутри блока (Split-Transform-Merge).	- Улучшенная выразительная способность сети без значительного увеличения числа параметров.- Высокая точность при умеренном росте вычислительной нагрузки.- Хорошо масштабируется для сложных задач.	- Более сложная архитектура, чем у ResNet.- Требуется аккуратная настройка гиперпараметров (кардинальность, ширина слоев).
EfficientNet	Использует compound scaling — одновременно масштабирует глубину, ширину и разрешение входного изображения.	- Эффективное использование ресурсов.- Меньшее количество параметров и FLOPs при сопоставимой точности.- Подходит для мобильных и встроенных устройств.	- Менее гибкая при необходимости точной настройки под отдельные задачи.- Для очень больших датасетов может уступать глубоким ResNet/ResNeXt по точности.

Таблица 3. Результаты классификации типов БПЛА

Архитектура	Точность	Скорость, с
ResNet-50	0,9038	0,19
ResNeXt-50	0,9135	0,30
EfficientNet	0,8710	0,17

Эксперименты по классификации проводились на датасете из 712 изображений, распределенных по шести классам. По совокупности характеристик наиболее сбалансированным выбором оказалась архитектура ResNet-50, тогда как ResNeXt-50 показала наивысшую точность — 0,9135. Наибольшие трудности возникали при различении классов hexacopter и fixed-wing drone из-за ограниченного количества соответствующих образцов в выборке.

Реализация системы

Подсистема классификации реализована не как отдельное приложение, а как составная часть общего программного комплекса, адаптированная под его архитектуру и требования. Основное приложение разработано на языке C# с использованием платформы .NET, поэтому интеграция модели машинного обучения также выполнена с учётом этой технологической базы. Такой подход обеспечивает единообразие среды выполнения, упрощает сопровождение системы и исключает необходимость использования сторонних интерпретируемых сред на этапе эксплуатации.

Обучение нейронной сети выполняется отдельно, в среде Python с использованием библиотеки PyTorch, что позволяет эффективно настраивать архитектуру (в частности, ResNet) и работать с обучающими данными. Однако итоговая модель не используется напрямую в Python-среде, а проходит этап адаптации для встраивания в основное приложение.

Ключевым этапом является преобразование обученной модели в формат ONNX, который выступает в роли универсального промежуточного представления. В процессе экспорта происходит восстановление архитектуры сети, загрузка обученных весов и формирование вычислительного графа, совместимого с платформой .NET. Такой подход позволяет перенести модель из исследовательской среды в промышленную без потери её функциональности.

Использование ONNX обеспечивает возможность выполнения инференса непосредственно в C#-приложении с помощью библиотеки ML.NET. Это означает, что классификация изображений выполняется внутри основного приложения, без вызова внешних сервисов или использования Python, что положительно сказывается на производительности и надёжности системы.

Таким образом, подсистема классификации представляет собой интегрированный компонент общего приложения, адаптированный под стек технологий C#/.NET и использующий формат ONNX для эффективного взаимодействия между этапами обучения и эксплуатации модели. Такой подход обеспечивает удобство внедрения, высокую скорость работы и возможность дальнейшего расширения функциональности системы.

Одной из важных функций интерфейса является выбор области интереса ROI на основном видеопотоке. Оператор может интерактивно выделить область, в пределах которой далее выполняется сопровождение цели. Это позволяет локализовать объект интереса, уменьшить влияние фоновых объектов и повысить устойчивость трекинга в реальных сценах. Дополнительно реализованы команды переключения между камерой и видеозаписью, запуска и остановки записи, сохранения видео и создания снимков. Эти средства делают интерфейс не только инструментом оперативного управления, но и средством последующего анализа результатов работы комплекса.

Управление аппаратными средствами комплекса интегрировано непосредственно в главное окно. Оператор получает доступ к подключению и настройке дальномера, заданию параметров двигателя, включению и выключению лазера, изменению мощности и контролю состояния устройств. Наличие единого интерфейса для наблюдения, сопровождения, наведения и фиксации результатов существенно снижает когнитивную нагрузку на оператора и создает предпосылки для практического применения системы в режиме реального времени.

Экспериментальные результаты и обсуждение

Эксперименты по обучению и тестированию моделей классификации проводились в среде Google Colab с использованием графического ускорителя NVIDIA Tesla T4. Это позволило обеспечить достаточную вычислительную мощность для обучения глубоких нейронных сетей и проведения серии сравнительных экспериментов в сопоставимых условиях. Обучение выполнялось на подготовленном датасете изображений, включающем различные классы объектов с учётом вариативности ракурсов, освещения и фона.

В рамках исследования были рассмотрены архитектуры ResNet-50, ResNeXt-50 и EfficientNet. Все модели обучались с использованием одинаковых параметров и обучающей выборки, что позволило объективно оценить их характеристики. Наилучшую точность классификации продемонстрировала модель ResNeXt-50, благодаря использованию параллельных преобразований внутри блоков сети. Однако увеличение сложности архитектуры привело к росту вычислительных затрат и времени инференса.

Модель EfficientNet показала хорошее соотношение точности и количества параметров, однако в процессе дообучения на относительно небольшом

пользовательском датасете её поведение оказалось менее стабильным. В частности, наблюдалась чувствительность к изменениям гиперпараметров и некоторая вариативность результатов на валидационной выборке.

Наиболее сбалансированные результаты продемонстрировала архитектура ResNet-50. Она обеспечила высокую точность классификации, стабильную сходимость в процессе обучения и предсказуемое поведение при изменении параметров. При этом модель показала меньшие требования к вычислительным ресурсам по сравнению с ResNeXt, что является важным фактором при использовании в системах реального времени.

С точки зрения практической реализации подсистема классификации была адаптирована под общее приложение, разработанное на C# и платформе .NET. После завершения обучения модель экспортировалась в формат ONNX, что позволило интегрировать её непосредственно в основное приложение без использования Python на этапе эксплуатации. Инференс выполняется с использованием ML.NET, обеспечивая высокую скорость обработки и стабильность работы.

В составе комплексной системы классификация функционирует как один из этапов общего вычислительного конвейера, дополняя задачи детекции, сопровождения и анализа движения объектов. Результаты классификации используются для уточнения типа цели и могут быть отображены в пользовательском интерфейсе, обеспечивая оператору дополнительную информацию для принятия решений.

Таким образом, проведённые эксперименты подтвердили целесообразность выбора архитектуры ResNet-50 в качестве основной модели классификации. Она обеспечивает оптимальное сочетание точности, скорости и устойчивости, а также хорошо интегрируется в программный комплекс благодаря использованию формата ONNX и средств платформы .NET.

Выводы

В рамках работы была рассмотрена и реализована подсистема классификации типов объектов в составе комплексной системы обнаружения и сопровождения БПЛА на основе методов компьютерного зрения. Проведён сравнительный анализ современных архитектур глубоких сверточных нейронных сетей, включая ResNet-50, ResNeXt-50 и EfficientNet, позволил определить наиболее эффективное решение с учётом требований прикладной эксплуатации.

Экспериментальные исследования показали, что архитектура ResNeXt-50 обеспечивает наивысшую точность классификации (до 91,35%), однако характеризуется повышенной вычислительной сложностью и менее стабильным временем инференса. EfficientNet продемонстрировала хорошее соотношение точности и количества параметров, но оказалась менее устойчивой при обучении и дообучении на ограниченном датасете.

Наиболее сбалансированным решением признана архитектура ResNet-50, которая обеспечила стабильную сходимость, высокую точность классификации и предсказуемое поведение при изменении параметров обучения. Дополнительным преимуществом данной модели является её простота и широкая поддержка, что облегчает её адаптацию и интеграцию в прикладные системы.

С практической точки зрения подсистема классификации была интегрирована в основное C#-приложение и адаптирована под архитектуру .NET. Для этого обученная модель была экспортирована в формат ONNX, что позволило выполнять инференс непосредственно в среде приложения с использованием ML.NET без необходимости обращения к Python на этапе эксплуатации. Такой подход обеспечил высокую производительность и упростил процесс внедрения модели в систему.

Таким образом, разработанная подсистема классификации демонстрирует корректную работу в составе общего программного комплекса и обеспечивает решение задачи определения типа объекта с достаточной точностью и стабильностью. Перспективы дальнейшего развития связаны с расширением обучающей выборки, повышением устойчивости моделей к сложным условиям съёмки и исследованием более лёгких архитектур для работы в реальном времени.

Работа выполнена в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования Российской Федерации (тема FZUN-2024-0020, госзадание ВлГУ).

Литература

1. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // CVPR. 2016.
2. He K., Zhang X., Ren S., Sun J. Identity Mappings in Deep Residual Networks // ECCV. 2016.
3. Бахвалов И.М., Старцев С.В. Применение глубоких сверточных нейронных сетей для задач распознавания изображений // Научный аспект. 2023.
4. Загоровский А.В., Киселёв Д.А. Использование остаточных нейронных сетей в задачах компьютерного зрения // Вестник МГТУ им. Баумана. 2022.
5. Кузнецов П.А., Лебедев И.С. Анализ эффективности ансамблевых сверточных архитектур в задачах классификации изображений // Информатика и системы управления. 2023.
6. Иванов С.В., Петров А.Н. Применение EfficientNet в задачах классификации изображений на ограниченных устройствах // Известия вузов. 2022.
7. Zhang H., Wu C., Zhang Z. ResNeXt for Image Classification: A Practical Study // IEEE Access. 2020.
8. Кузнецов П.А., Лебедев И.С. Анализ эффективности ансамблевых сверточных архитектур в задачах классификации изображений // Информатика и системы управления. 2023.
9. Li Y., Wang N., Shi J. Revisiting ResNeXt for Efficient Visual Recognition // Pattern Recognition Letters. 2021.