М

УДК 004.932.2 **DOI** 10.24412/2413-7383-2025-4-39-60-70

Е. С. Мороз, Я. С. Пикалёв
Федеральное государственное бюджетное научное учреждение
«Институт проблем искусственного интеллекта», г. Донецк
283048, г. Донецк, ул. Артема, 118 б

АНАЛИЗ Sota-АРХИТЕКТУР ДЛЯ СЕМАНТИЧЕСКОЙ АУГМЕНТАЦИИ*

Ye. S. Moroz, Y.S. Pikaliov
Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»
283048, Donetsk, Artema str, 118-b.

ANALYSIS OF SOTA-ARCHITECTURES FOR SEMANTIC AUGMENTATION

В статье рассмотрены задачи семантической сегментации в контексте прикладного компьютерного зрения и связанного круга задач: обнаружение объектов, сегментация экземпляров и панорамная сегментация. Проанализированы лёгкие и высокоскоростные архитектуры реального времени, с точки зрения баланса между точностью и задержкой на наборах Cityscapes, CamVid и UAVid. Основное внимание уделено вопросам подготовки данных: аугментации, аннотированию изображений, предварительной обработке и интерпретируемости моделей.
Ключевые слова: семантическая сегментация, компьютерное зрение, обнаружение объектов, сегментация экземпляров, панорамная сегментация, аугментация данных, перекрёстная геолокализация.

The paper considers semantic segmentation tasks in the context of applied computer vision and related problems such as object detection, instance segmentation, and panoptic segmentation. Lightweight and high-speed real-time architectures are analyzed in terms of the trade-off between accuracy and latency on the Cityscapes, CamVid, and UAVid datasets. Special attention is paid to data preparation issues, including augmentation, image annotation, preprocessing, and model interpretability.
Key words: semantic segmentation, computer vision, object detection, instance segmentation, panoptic segmentation, data augmentation, cross-view geolocalization.

* Статья выполнена в Федерального государственного бюджетного научного учреждения «Институт проблем искусственного интеллекта». «Лаборатории интеллектуальных систем и анализа данных»

Введение

Семантическая сегментация в последние годы приобрела особую значимость как ключевая задача компьютерного зрения, направленная на классификацию каждого пикселя изображения [1]. Этот метод позволяет системам интерпретировать сцены, выделяя объекты различных классов и формируя карту визуальной информации. С развитием глубокого обучения и ростом объёмов данных её эффективность существенно возросла, в том числе при работе с искусственными цифровыми изображениями [2].

Цель работы – оценка подхода к применению семантической сегментации аэрофотоснимков для задач перекрёстной геолокации с использованием аугментации данных. В настоящий момент семантическая сегментация применяется в автономном вождении, медицинской диагностике, робототехнике, промышленном контроле, анализе окружающей среды, а также при навигации по аэрофотоснимкам [3]. Она используется для обнаружения ключевых объектов и уточнения их пространственного положения на снимках местности, что требует специализированных методик работы с аэрофотоснимками [4] и продуманного формирования целевых наборов изображений [5]. При этом качество разметки и устойчивость моделей существенно зависят от выбора методов аугментации, расширяющих обучающие выборки и повышающих вариативность данных [6].

Постановка задачи исследования заключается в том, чтобы по результатам анализа рассмотренных SOTA-архитектур лёгкой семантической сегментации оценить их пригодность для обработки аугментированных наборов аэрофотоснимков (UAVid, Aeroscapes, iSAID, xView, Drone Tracking, SynDrone), принимая во внимание, что аугментация изменяет распределение обучающих данных и усложняет их вариативность, что может влиять как на точность и устойчивость сегментации, так и на допустимые вычислительные затраты при выборе архитектуры.

Основная часть

Для лучшего понимания роли семантической сегментации рассмотрим связанные задачи компьютерного зрения, направленные на анализ содержания изображений, а также вопросы подготовки и аннотирования данных для обучения соответствующих моделей [7].

1. Обнаружение объектов (*Object Detection*) - ключевая задача, направленная на поиск, локализацию и классификацию объектов на изображении. Цель - определить положение каждого объекта с помощью ограничивающей рамки (bounding box) и отнести его к определённому классу. Для успешного решения задачи важно обеспечить корректную предварительную обработку изображений и формирование признаков, удобных для последующего обучения [8].

2. Ориентированное обнаружение (*Oriented Object Detection*) расширяет классическую детекцию, учитывая ориентацию объектов и их форму. Вместо осево-ориентированных рамок применяются рамки с произвольным углом наклона, что позволяет точнее описывать наклонённые и продолговатые объекты. Преимущества заключаются в более точной локализации объектов произвольной ориентации и формы, а также в лучшей адаптации к реальным сценам, включая видеопоследовательности и сложные динамические сюжеты [9]. Недостатком является увеличение вычислительной сложности и необходимость специальных алгоритмов для оценки ориентации.

3. Сегментация экземпляров (*Instance Segmentation*) направлена на выделение и классификацию отдельных объектов на изображении, в отличие от семантической

сегментации, которая группирует пиксели по классам без разделения разных объектов одного типа. Она позволяет отличать отдельные объекты друг от друга, что важно для робототехники, медицинской диагностики и автономных систем. Для повышения качества таких моделей применяются расширенные техники аугментации и построения наборов данных изображений [10].

4. Семантическая сегментация (*Semantic Segmentation*) решает задачу классификации каждого пикселя изображения - результатом является полноценная карта, где каждому пикселю присвоен класс из заданного множества. Этот подход обеспечивает полное понимание сцены на уровне пикселей и находит применение там, где требуется детальный анализ: диагностика по изображениям, оценка состояния объектов и интерпретация пространственных структур, что ставит задачу интерпретируемости нейросемантических моделей в прикладных областях [11].

5. Панорамная сегментация (*Panoptic Segmentation*) классифицирует каждое пиксельное значение и разделяет объекты по экземплярам, обеспечивая полное понимание сцены, включая как фоновые классы, так и конкретные объекты. Методы панорамной сегментации, как правило, используют сложные мультизадачные архитектуры, например объединение выходов семантической и экземплярной сегментации, и предъявляют высокие требования к качеству разметки и устойчивости моделей, что подтверждается исследованиями инстанс-сегментации сложных объектов на реальных изображениях [12].

Развитие приложений, требующих обработки видео и изображений в реальном времени (например, на борту беспилотника, в автомобиле или на мобильном устройстве), стимулировало создание специализированных облегчённых и высокоскоростных архитектур сегментации. Ниже рассмотрены современные модели, ориентированные на работу в реальном времени при сохранении высокой точности.

1. BiSeNet - архитектура семантической сегментации для задач компьютерного зрения (computer vision). Решение использует двуветвевую схему: пространственный путь (spatial path) с малым шагом свёрток формирует представления высокого разрешения, сохраняя геометрию и границы, а контекстный путь (context path) с быстрой понижающей дискретизацией расширяет поле восприятия. В контекстном пути задействован модуль уточнения внимания (attention refinement module) и глобальное усреднение по пространству, что стабилизирует агрегирование признаков крупного масштаба. Слияние ветвей выполняет модуль слияния признаков (feature fusion module), который адаптивно объединяет детальные и контекстные признаки и балансирует вклад каналов. Подход реализуется на стандартных лёгких CNN-бэкбонах, допускает переключение разрешения входа и сохраняет работу в реальном времени. На наборе Cityscapes для конфигурации с Xception (бэкбон) выдаёт 68,4% mIoU при 105,8 кадрах в секунду на NVIDIA Titan Xp при входе 1536x768, при полном 2048x1024 скорость ниже при сопоставимой точности.

Преимуществами BiSeNet являются: сохранение мелких деталей за счёт пространственного пути, устойчивый захват контекста при помощи контекстного пути и модуль уточнения внимания, высокая скорость инференса (вывод результатов работы обученной модели, в процессе решения задачи на новых данных) на практических разрешениях. Недостатками являются: ограничение верхней точности относительно современных трансформеров, чувствительность к настройке слияния признаков в блоке ffm, снижение качества на редких и мелких классах без доменной донастройки [13].

2. PIDNet-S. Система реализует трёхветвевую схему: P-ветвь обрабатывает высокое разрешение и сохраняет мелкие структуры, I-ветвь быстро понижает разрешение и накапливает контекст, D-ветвь выделяет границы и по карте границ направляет слияние ветвей деталей и контекста в узле, направляя объединение P и I для повышения точности на границах объектов; лёгкие блоки и взвешивание каналов уменьшают смешивание классов на стыках и сохраняют высокую частоту кадров. На наборе Cityscapes точность по метрике mIoU равен 78,6% при 93,2 кадра в секунду, на CamVid - 80,1% mIoU при 153,7 кадра в секунду, что подтверждает компромисс между скоростью и точностью для применения в реальном времени [14]. Решение ориентировано на сцены с выраженными границами и тонкими объектами, где качество контуров критично для последующей постобработки и трекинга.

Преимуществами PIDNet-S являются: высокая чёткость контуров за счёт граничного внимания, конкурентный баланс скорости и точности на городских сценах, сохранение деталей при работе в реальном времени. Недостатками являются: ограничение верхней точности относительно тяжёлых трансформерных схем, чувствительность к качеству разметки границ и настройкам слияния, снижение устойчивости на редких и мелких классах без доменной донастройки.

3. DDRNet. Сеть с ветвями высокого и низкого разрешения, между которыми выполняются многократные билатеральные слияния. Ветвь низкого разрешения ускоряет накопление контекста за счёт глубокой понижающей дискретизации, а ветвь высокого разрешения сохраняет геометрию и границы объектов. Модуль пирамидального пулинга с глубокой агрегацией (deep aggregation pyramid pooling module) расширяет эффективное поле восприятия и агрегирует многошкальный контекст на картах низкого разрешения. Билатеральные слияния и DAPPM совместно обеспечивают согласование детальных и контекстных представлений при ограниченных вычислительных затратах.

На наборе Cityscapes точность равна 77,4% mIoU при 102 кадрах в секунду на RTX 2080 Ti, на наборе CamVid - 74,7% mIoU при 230 кадрах в секунду, что отражает компромисс между точностью и скоростью [15]. Преимуществами DDRNet являются: высокая чёткость границ за счёт ветви высокого разрешения, сбор контекста с разных масштабов блоком DAPPM, конкурентная скорость инференса. Недостатками являются: ограничение верхней точности относительно тяжёлых трансформерных схем, чувствительность к балансу между ветвями, необходимость доменной донастройки при переносе на новые сцены.

4. SegFormer. Модель с иерархическим энкодером и лёгким декодером – извлекает признаки на разных масштабах и учитывает дальние связи. Энкодер делит изображение на пересекающиеся куски и обрабатывает их слоями, что позволяет отказаться от позиционного кодирования без потери локальной привязки. Декодер агрегирует представления из нескольких уровней и восстанавливает карту классов без тяжёлых операций апсемплинга, что упрощает инференс.

Конфигурации, такие как B0 ориентированы на малое число параметров и низкую задержку инференса при сохранении качества на городских и аэровизуальных сценах. На наборе UAVid точность mIoU равна 66,19% при 147 кадрах в секунду, что допускает выполнение на борту при ограниченных вычислительных ресурсах [16].

Преимуществами SegFormer-B0 являются: низкая задержка инференса, компактность модели, устойчивый захват контекста и мелких деталей. Недостатками являются: ограничение верхней точности относительно крупных конфигураций SegFormer, зависимость метрик от входного разрешения и аппаратной платформы, необходимость доменной донастройки для редких и мелких классов.

5. DF1-Seg - средство семантической сегментации для задач компьютерного зрения. Решение опирается на алгоритм частичного упорядоченного отсечения (partial order pruning), который применяется для поиска архитектуры декодера и формирует семейство сетей, ориентированных на работу в реальном времени. Поиск декодера упрощает процесс восстановления пространственного разрешения и снижает вычислительные затраты без усложнения апсемплинга. Бэкбоны DF обеспечивают высокую пропускную способность и совместимы со стандартными схемами обучения, что позволяет масштабировать вход и управлять компромиссом между скоростью и точностью. По результатам классификационных испытаний DF1 и DF2 достигают точности выше, чем ResNet-18 и ResNet-50, при меньшей латентности на сопоставимом оборудовании [17]. В конфигурации сегментации DF1-Seg демонстрирует работу в реальном времени как на производительных графических процессорах, так и на встраиваемых системах. На наборе Cityscapes mIoU равно 74,1% при 106,4 кадра в секунду на GTX 1080 Ti при разрешении 1024x2048, что подтверждает пригодность для покадровой обработки городских сцен. На Jetson TX2 достигается 21,8 кадра в секунду при 1280x720 при сохранении стабильности инференса и умеренного потребления памяти. Архитектура допускает доменную донастройку на целевых данных и совместима с приёмами ускорения инференса за счёт понижения точности вычислений.

Преимуществами DF1-Seg являются: выраженный баланс скорости и точности на разных классах устройств, компактный декодер из поиска POP, низкая латентность на встраиваемых платформах, предсказуемое масштабирование по разрешению. Недостатками являются: ограничение верхней точности относительно тяжёлых трансформерных схем, чувствительность к выбору бэкбона и параметров поиска, необходимость доменной донастройки при переносе на новые сцены.

6. SFNet. Решение реализует модуль выравнивания потока (flow alignment module), который оценивает семантический поток и выравнивает многоуровневые признаки перед слиянием. Такое выравнивание корректирует расхождения между масштабами и повышает качество границ без существенного вычисления. Система поддерживает разные бэкбоны, что позволяет выбирать компромисс между глубиной и скоростью под целевое оборудование. В конфигурации с ResNet-18 на наборе Cityscapes mIoU равно 80,4% при 26 кадрах в секунду на GTX 1080 Ti при предварительном предобучении на больших уличных датасетах с пиксельной разметкой (Mapillary Vistas), без предобучения - 78,9% mIoU. При использовании DF2 в качестве бэкбона достигается 77,8% mIoU при 61 кадре в секунду и 74,5% mIoU при 121 кадре в секунду в разных режимах скорости и точности. С более глубокими бэкбонами, например ResNet-101, отмечается до 81,8% mIoU, при этом итоговая латентность зависит от стека инференса и входного разрешения [18].

Преимуществами SFNet являются: точное слияние многоуровневых признаков за счёт модуля выравнивания потока, масштабируемости по бэкбонам, конкурентный баланс скорости и точности, прирост от предобучения на уличных датасетах. Недостатками являются: чувствительность к выбору бэкбона и режиму инференса, снижение показателей без предобучения и оптимизаций, необходимость доменной донастройки при переносе на новые сцены.

7. BiSeNet v2. Решение реализует двуветвевую схему без трансформеров: ветвь детализации (detail branch) с широкими каналами и малой глубиной формирует представления высокого разрешения, а семантическая ветвь (semantic branch) с узкими каналами и глубокой иерархией накапливает контекст при понижении разрешения.

В семантической ветви вычислительная нагрузка снижается за счёт быстрой стратегии уменьшения разрешения и лёгких блоков на по-канальных свёртках (depthwise), дополненных модулем контекстного встраивания (context embedding), который стабилизирует учёт глобального контекста. Двусторонний направляемый слой агрегации (bilateral guided aggregation layer) обеспечивает направляемое объединение детальных и контекстных признаков и уменьшает смешивание классов на границах. В обучении применяется схема с дополнительными головами, которые удаляются на инференсе, что повышает точность без накладных расходов в рабочем режиме.

На наборе Cityscapes при входе 2048x1024 точность равна 72,6% mIoU при 156 кадрах в секунду на NVIDIA GTX 1080 Ti, что подтверждает пригодность для обработки видео в реальном времени [19]. Архитектура ориентирована на сцены с выраженными границами и небольшой толщиной объектов, где важны сохранение деталей и устойчивый контекст, и позволяет балансировать скорость и качество через выбор разрешения и глубины бэкбона. Преимуществами BiSeNet v2 являются: высокая скорость инференса при полном входном разрешении, раздельная обработка деталей и контекста через DB и SB, отсутствие накладных расходов от обучения на стадии работы. Недостатками являются: ограничение верхней точности относительно тяжёлых трансформерных схем, чувствительность к качеству слияния признаков на границах, необходимость доменной донастройки для редких и мелких классов.

8. FasterSeg. Архитектуру подобрали автопоиском в пространстве с ветвями разных разрешений, чтобы учитывать и детали, и общий контекст. В процессе оптимизации применяется раздельная и детализированная регуляризация задержки, которая удерживает поиск от смещения к вариантам с низкой задержкой и низкой точностью. Система поддерживает совместный поиск с дистилляцией знаний от учителя к ученику, что повышает итоговую точность без утяжеления тракта инференса. Поиск ориентируется на оценку задержки для целевого устройства, что делает итоговую конфигурацию чувствительной к профилю устройства, но обеспечивает предсказуемую скорость.

На наборе Cityscapes при разрешении 1024x2048 точность достигает 71,5% mIoU при 163,9 кадра в секунду на NVIDIA GTX 1080 Ti, что соответствует применению в реальном времени [20]. Численные значения зависят от входного разрешения, движка инференса и особенностей предобработки. Архитектура допускает масштабирование по разрешению и числу каналов и интегрируется с приёмами ускорения, при сохранении общей структуры. Преимуществами FasterSeg являются: многомасштабные ветки, найденные автопоиском, и раздельный контроль задержки по блокам модели, прирост качества от совместного поиска с переносом знаний, высокая кадровая частота на Cityscapes. Недостатками являются: зависимость результата от профилирования конкретного GPU, чувствительность к настройкам поиска и прокси латентности, ограничение верхней точности относительно тяжёлых трансформерных схем, необходимость доменной донастройки при переносе.

9. TinyHMSeg. Модель использует модуль лёгкого контекстного слияния (Lightweight Context Fusion), который выравнивает и объединяет признаки на разных масштабах перед восстановлением карты классов. Лёгкий модуль глобального усиления (Lightweight Global Enhancement module) собирает общий контекст кадра и настраивает веса каналов, формируя небольшой выходной блок сегментации без тяжёлого декодера. Структура с несколькими уровнями разрешения в бэкбоне балансирует размер карт и число каналов, что снижает вычислительные затраты при сохранении устойчивых признаков. Такая схема сети позволяет удерживать низкую задержку и предсказуемый расход памяти на дискретных и встраиваемых графических процессорах.

На наборе Cityscapes точность достигает 71,4% mIoU при 172,4 кадра в секунду на NVIDIA GTX 1080 Ti, что указывает на пригодность для потоковой обработки в реальном времени [21]. Подход применим для покадровой обработки видео с фиксацией ключевых кадров в последующей постобработке и допускает доменную донастройку на целевых данных. Преимуществами TinyHMSeg являются: высокая скорость инференса при сопоставимой точности, компактная голова сегментации за счёт LCF и LGE, стабильное масштабирование по разрешению и памяти. Недостатками являются: снижение точности на границах и мелких объектах при низком внутреннем разрешении, зависимость метрик от оптимизаций инференса и предобучения, необходимость доменной донастройки при переносе на иные сцены.

10. STDC-Seg. Решение пересматривает двуветвевую идею BiSeNet, устраняя структурную избыточность за счёт сокращения каналов в ранних слоях и упрощения тракта декодирования. Базовый блок краткосрочной плотной конкатенации (short-term dense concatenate) формирует сжатые признаки, объединяя карты признаков соседних ступеней с последующим сведением каналов. В декодере предусмотрен модуль объединения деталей, который однопоточно вводит пространственную информацию в низкоуровневые слои. Далее низкоуровневые и высокоуровневые признаки объединяются для предсказания карты классов, что обеспечивает согласование границ и контекста при малых вычислительных затратах. На наборе Cityscapes точность достигает 71,9% mIoU при 250,4 кадра в секунду на GTX 1080 Ti, при повышенном входном разрешении достигается 76,8% mIoU при 97,0 кадра в секунду, что отражает масштабируемость по качеству и скорости [22]. Преимуществами STDC-Seg являются: высокая скорость инференса на практических разрешениях, компактные представления за счёт блока STDC, согласование деталей и контекста в простом декодере. Недостатками являются: ограничение верхней точности относительно трансформерных схем, чувствительность к выбору разрешения и степени сжатия каналов, необходимость доменной донастройки для редких и мелких классов.

Заключение

Сравнительный анализ показал, что среди современных моделей реального времени нет универсального решения, и каждая архитектура занимает свою нишу в зависимости от требований к скорости, качеству и ресурсам. Сети семейства BiSeNet и BiSeNet v2 ориентированы на высокую частоту кадров при полном входном разрешении, что делает их удобными для задач, где важна интерактивность и непрерывная потоковая обработка, например для мобильных и робототехнических платформ в городских сценах. PIDNet-S и SFNet демонстрируют повышенную точность на границах и тонких объектах и более всего подходят для сценариев, где критична детализация контуров дорожной разметки, инженерных объектов и мелких элементов инфраструктуры. DDRNet и DF1-Seg обеспечивают сбалансированный компромисс между точностью и задержкой на стандартных графических процессорах и могут рассматриваться как базовые варианты для систем дорожного анализа и промышленного мониторинга, где допустим средний уровень вычислительной нагрузки при высокой стабильности работы. SegFormer-B0 показывает целесообразность использования трансформерных подходов в компактных конфигурациях и особенно интересен для аэровизуальных наборов данных и задач бортовой обработки, где требуется широкий контекст при ограниченных ресурсах. TinyHMSeg и FasterSeg ориентированы на жёсткие ограничения по памяти и времени отклика, а потому целесообразны при развёртывании на встраиваемых платформах и специализированных устройствах, в том

числе с учётом аппаратно-ориентированного поиска архитектуры. STDC-Seg демонстрирует высокую скорость при сохранении конкурентного качества и подходит для систем, где важно масштабировать разрешение под конкретный режим эксплуатации без перепроектирования всей модели.

В совокупности рассмотренные решения формируют набор типовых архитектурных шаблонов, позволяющих подбирать модель под конкретный класс устройств, характеристики данных и допустимую задержку.

Список литературы

1. Мороз Е. С. Методы семантической сегментации в компьютерном зрении: обзор архитектур, функций потерь и современных подходов // Донецкие чтения – 2025: образование, наука, инновации, культура и вызовы современности : материалы X Международной научной конференции (Донецк, 2025 г.). Донецк, 2025.
2. Покинтелица А. Е. Применение клеточных автоматов для создания искусственных цифровых изображений // Искусственный интеллект: теоретические аспекты и практическое применение: материалы Донецкого Международного круглого стола. Донецк : ФГБНУ «ИПИИ», 2025. 296 с. С. 12–18.
3. Ермоленко Т. В. К вопросу о применении глубокого обучения для задачи перекрёстной геолокализации / Т. В. Ермоленко, Р. С. Хакимов // Проблемы искусственного интеллекта. 2024. № 4 (35). С. 12–28.
4. Пикалёв, Я. С. Обнаружение ключевых объектов и перекрёстная геолокация: анализ наборов данных и методологические аспекты // Проблемы искусственного интеллекта. 2024. № 4.С. 25–37. ISSN 2413-7383.
5. Павленко, Б. В. Методика создания набора аэрофотоснимков для задачи перекрёстной геолокации / Б. В. Павленко, Я. С. Пикалёв // Проблемы искусственного интеллекта. 2024. № 4. С. 101–112. ISSN 2413-7383.
6. Близно М. В. Сравнение различных методов аугментации на обучение модели нейронной сети // Искусственный интеллект: теоретические аспекты и практическое применение: материалы Донецкого Международного круглого стола. Донецк : ФГБНУ «ИПИИ», 2025. 296 с. С. 99–102.
7. Хакимов Р. С. К вопросу о разработке системы аннотирования данных для задач компьютерного зрения / Р. С. Хакимов, О. Л. Нижникова, М. В. Близно // Проблемы искусственного интеллекта. 2024. № 3 (34). С. 70–79. ISSN 2413-7383. DOI 10.24412/2413-7383-2024-3-70-79.
8. Акимов А. А. Предварительная обработка данных для машинного обучения / А. А. Акимов, Д. Р. Валитов, А. И. Кубряк // Научное обозрение. Технические науки. 2022. № 2. С. 26–31.
9. Кришнан Ш. Р. Улучшение обнаружения аномалий на видео с помощью усовершенствованной технологии UNET и техники каскадного скользящего окна / Ш. Р. Кришнан, П. Амудха // Информатика и автоматизация. 2024. № 6 (23). С. 1899–1930. ISSN 2713-3192. – DOI 10.15622/ia.23.6.12.
10. Хакимов Р. С. Обзор расширенных техник аугментации для набора данных изображений / Р. С. Хакимов, Б. В. Павленко, Я. С. Пикалёв // Донецкие чтения 2024: материалы IX Международной научной конференции (Донецк, 15–17 октября 2024 г.). Т. 2. Донецк : Изд-во ДонГУ, 2024. С. 272–275. ISSN 2664-7362; ISSN 2664-7370.
11. Никитенко, К. А. Интерпретируемость нейросемантических моделей при их применении в прикладных областях / К. А. Никитенко, А. В. Звягинцева. // Проблемы искусственного интеллекта. 2025. № 2. С. 79–90. ISSN 2413-7383.
12. Солопов, М. В. Исследование применения моделей Mask R-CNN и Segment Anything Model (SAM) для инстанс-сегментации мезенхимных стволовых клеток на микрофотографиях / М. В. Солопов, Е. С. Чечехина, А. Г. Попандопуло, А. С. Кавелина, Г. В. Акопян, В. В. Турчин // Проблемы искусственного интеллекта. 2025. № 2(37). С. 21–29. DOI 10.24412/2413-7383-2025-2-37-21-29. EDN ODOKND. ISSN 2413-7383.
13. Yu C. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation / C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang // Proc. ECCV. 2018. Electronic resource. Mode of access: https://www.alphaxiv.org/abs/1808.00897
14. Xu J. PIDNet: A Real-time Semantic Segmentation Network Inspired by PID Controllers / J. Xu, Y. Tang, K. Chen [et al.]. – arXiv preprint, 2022. – Electronic resource. – Mode of access: https://arxiv.org/abs/2206.02066
15. Yin H. Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes / H. Yin, X. Shen, J. Fang [et al.]. arXiv preprint, 2021. Electronic resource. Mode of access: https://arxiv.org/pdf/2101.06085
16. Xie E. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / E. Xie, W. Wang, Z. Yu [et al.]. arXiv preprint, 2024. Electronic resource. – Mode of access: https://arxiv.org/abs/2410.01092v1

17. Li H. DFNet: Deep Feature Aggregation Network for Real-time Semantic Segmentation / H. Li, P. Xiong, J. Fan, L. Sun. – arXiv preprint, 2019. – Electronic resource. – Mode of access: https://arxiv.org/abs/1903.03777v2
18. Li X. Semantic Flow for Fast and Accurate Scene Parsing (SFNet) / X. Li, H. Xiong, H. Fan [et al.] // Proc. ECCV. 2020. Electronic resource. Mode of access: https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123460749.pdf
19. Yu C. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation / C. Yu, C. Gao, J. Wang, G. Yu, N. Sang. arXiv preprint, 2020. Electronic resource. Mode of access: https://arxiv.org/pdf/2004.02147
20. Chen X. FasterSeg: Searching for Faster Real-time Semantic Segmentation / X. Chen, Y. Zhu, G. Lin [et al.]. – arXiv preprint, 2019. – Electronic resource. – Mode of access: https://arxiv.org/abs/1912.10917
21. When Humans Meet Machines: Towards Efficient Segmentation. – [s.l.], 2023. – Electronic resource. – Mode of access: https://scispace.com/pdf/when-humans-meet-machines-towards-efficient-segmentation-4pir2uvpjc.pdf
22. Fan M. Rethinking BiSeNet for Real-time Semantic Segmentation (STDC-Seg) / M. Fan, T. Wang, D. Gong, J. Yang. arXiv preprint, 2021. Electronic resource. – Mode of access: https://arxiv.org/abs/2104.13188

References

1. Moroz YE. S. Methods of semantic segmentation in computer vision: a review of architectures, loss functions and modern approaches / E. S. Moroz // Donetsk Readings – 2025: education, science, innovation, culture and challenges of the present: materials of the 10th International scientific conference (Donetsk, 2025). – Donetsk, 2025.
2. Pokintelitsa A. E. Application of cellular automata for creating artificial digital images / A. E. Pokintelitsa // Artificial Intelligence: theoretical aspects and practical application: proceedings of the Donetsk International Round Table. – Donetsk: FSBSI “Institute for Problems of Artificial Intelligence”, 2025. – 296 p. – P. 12–18.
3. Ermolienko T. V. On the application of deep learning to the problem of cross-view geolocalization / T. V. Ermolienko, R. S. Khakimov // Problems of Artificial Intelligence. – 2024. – No. 4 (35). – P. 12–28.
4. Pikalev, Ya. S. Detection of key objects and cross-view geolocation: analysis of datasets and methodological aspects / Ya. S. Pikalev. // Problems of Artificial Intelligence. – 2024. – No. 4. – P. 25–37. – ISSN 2413-7383.
5. Pavlenko, B. V.; Pikalev, Ya. S. Methodology for creating a set of aerial images for the cross-view geolocation problem / B. V. Pavlenko, Ya. S. Pikalev. // Problems of Artificial Intelligence. – 2024. – No. 4. – P. 101–112. – ISSN 2413-7383.
6. Blizno M. V. Comparison of various augmentation methods for training a neural network model / M. V. Blizno // Artificial Intelligence: theoretical aspects and practical application: proceedings of the Donetsk International Round Table. – Donetsk: FSBSI “Institute for Problems of Artificial Intelligence”, 2025. – 296 p. – P. 99–102.
7. Khakimov R. S. On the development of a data annotation system for computer vision tasks / R. S. Khakimov, O. L. Nizhnikova, M. V. Blizno // Problems of Artificial Intelligence. – 2024. – No. 3 (34). – P. 70–79. – ISSN 2413-7383. – DOI 10.24412/2413-7383-2024-3-70-79.
8. Akimov A. A. Preprocessing of data for machine learning / A. A. Akimov, D. R. Valitov, A. I. Kubryak //Scientific Review. Engineering Sciences. – 2022. – No. 2. – P. 26–31.
9. Krishnan S. R. Improving anomaly detection in video using an enhanced UNET technology and a cascaded sliding-window technique / S. R. Krishnan, P. Amudha // Informatics and Automation. – 2024. – No. 6 (23). – P. 1899–1930. – ISSN 2713-3192. – DOI 10.15622/ia.23.6.12.
10. Khakimov R. S. Review of advanced augmentation techniques for image datasets / R. S. Khakimov, B. V. Pavlenko, Ya. S. Pikalev // Donetsk Readings 2024: materials of the 9th International scientific conference (Donetsk, 15–17 October 2024). – Vol. 2. – Donetsk: DonNU Press, 2024. – P. 272–275. – ISSN 2664-7362 (Print); ISSN 2664-7370.
11. Nikitenko, K. A.; Zvyagintseva, A. V. Interpretability of neurosemantic models in their application to practical domains / K. A. Nikitenko, A. V. Zvyagintseva. // Problems of Artificial Intelligence. – 2025. – No. 2. – P. 79–90. – ISSN 2413-7383.
12. Solopov, M. V.; Chechekhina, E. S.; Popandopulo, A. G.; Kavelina, A. S.; Akopyan, G. V.; Turchin, V. V. Study of the application of Mask R-CNN and Segment Anything Model (SAM) for instance segmentation of mesenchymal stem cells in microphotographs / M. V. Solopov, E. S. Chechekhina, A. G. Popandopulo,

A. S. Kavelina, G. V. Akopyan, V. V. Turchin. // Problems of Artificial Intelligence. – 2025. – No. 2(37). – P. 21–29. – DOI: 10.24412/2413-7383-2025-2-37-21-29. – EDN: ODOKND. – ISSN 2413-7383.
13. Yu C. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation / C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang // Proc. ECCV. – 2018. – Electronic resource. – Mode of access: https://www.alphaxiv.org/abs/1808.00897
14. Xu J. PIDNet: A Real-time Semantic Segmentation Network Inspired by PID Controllers / J. Xu, Y. Tang, K. Chen [et al.]. arXiv preprint, 2022. Electronic resource. Mode of access: https://arxiv.org/abs/2206.02066
15. Yin H. Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes / H. Yin, X. Shen, J. Fang [et al.]. arXiv preprint, 2021. Electronic resource. Mode of access: https://arxiv.org/pdf/2101.06085
16. Xie E. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / E. Xie, W. Wang, Z. Yu [et al.]. arXiv preprint, 2024. Electronic resource. Mode of access: https://arxiv.org/abs/2410.01092v1
17. Li H. DFNet: Deep Feature Aggregation Network for Real-time Semantic Segmentation / H. Li, P. Xiong, J. Fan, L. Sun. – arXiv preprint, 2019. – Electronic resource. – Mode of access: https://arxiv.org/abs/1903.03777v2
18. Li X. Semantic Flow for Fast and Accurate Scene Parsing (SFNet) / X. Li, H. Xiong, H. Fan [et al.] // Proc. ECCV. – 2020. – Electronic resource. – Mode of access: https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123460749.pdf
19. Yu C. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation / C. Yu, C. Gao, J. Wang, G. Yu, N. Sang. – arXiv preprint, 2020. – Electronic resource. – Mode of access: https://arxiv.org/pdf/2004.02147
20. Chen X. FasterSeg: Searching for Faster Real-time Semantic Segmentation / X. Chen, Y. Zhu, G. Lin [et al.]. – arXiv preprint, 2019. – Electronic resource. – Mode of access: https://arxiv.org/abs/1912.10917
21. When Humans Meet Machines: Towards Efficient Segmentation / [no authors]. – [s.l.], 2023. – Electronic resource. – Mode of access: https://scispace.com/pdf/when-humans-meet-machines-towards-efficient-segmentation-4pir2uvpjc.pdf
22. Fan M. Rethinking BiSeNet for Real-time Semantic Segmentation (STDC-Seg) / M. Fan, T. Wang, D. Gong, J. Yang. – arXiv preprint, 2021. – Electronic resource. – Mode of access: https://arxiv.org/abs/2104.13188

RESUME

YE. S. Moroz, Y.S. Pikaliov

Analysis of Sota-Architectures for Semantic Augmentation

Background: The growth of real-time computer vision tasks under strict latency and resource constraints increases the demand for lightweight semantic segmentation architectures. It is important to compare on-board and embedded models, identify common structural principles, and relate them to the full processing pipeline – from data preparation and annotation to post-processing and the use of segmentation maps in navigation and control.

Materials and methods: A review and comparative analysis of lightweight real-time semantic segmentation networks was carried out.

Results: The analysis showed that lightweight real-time semantic segmentation architectures are built according to similar schemes with separate detail and context branches, multiscale feature fusion, and simplified decoders.

Conclusion: The proposed models can be used when choosing the architecture of segmentation networks in applied computer vision systems for a given class of devices and operating mode.

РЕЗЮМЕ

Е. С. Мороз, Я. С. Пикалёв

Анализ Sota-архитектур для семантической аугментации

Актуальность: рост задач компьютерного зрения реального времени при жёстких ограничениях по задержке и ресурсам усиливает требование к лёгким архитектурам семантической сегментации; важно сопоставлять бортовые и встраиваемые модели, выделять общие структурные принципы и увязывать их с полным циклом обработки - от подготовки и аннотирования данных до постобработки и использования карт сегментации в навигации и контроле.

Материалы и методы: выполнен обзор и сравнительный анализ по лёгким сетям семантической сегментации реального времени;

Результаты: анализ показал, что лёгкие архитектуры семантической сегментации реального времени строятся по сходным схемам с разделением ветвей деталей и контекста, многошкальным объединением признаков и упрощёнными декодерами.

Вывод: предложенные модели могут использоваться при выборе архитектуры сегментационных сетей в прикладных системах компьютерного зрения под заданный класс устройств и режим работы.

Мороз Егор Сергеевич – инженер-исследователь, Федеральное государственное бюджетное научное учреждение, «Институт проблем искусственного интеллекта».

Область научных интересов: анализ данных, распознавание образов, компьютерное зрение, эл. почта azidaan.moroz@yandex.com, адрес: г. Донецк, ул. Егорова, д. 16, кв. 1, телефон: +7(949) 381-67-00.

Пикалёв Ярослав Сергеевич – кандидат техн. наук, старший научный сотрудник лаборатории интеллектуальных систем и анализа данных Федерального государственного бюджетного научного учреждения «Институт проблем искусственного интеллекта»,

Область научных интересов: цифровая обработка сигналов, анализ данных, распознавание образов, обработка естественного языка, компьютерное зрение, машинное обучение, нейронные сети, г. Донецк, эл. почта: i@pikaliov.ru, адрес: 283085, ДНР, г. Донецк, ул. Отважных, д. 19, кв. 85, телефон: +7 (949) 428-73-88.

Статья поступила в редакцию 30.09.2025