

Особенности применения нейронных сетей при распознавании антропогенных объектов с вариативными контурами.

Саидов А.Г., Жуков Д.В.¹

¹ВКА им. А.Ф. Можайского, Россия, Санкт-Петербург

Аннотация: В работе анализируется результативность методов семантической сегментации и сегментации экземпляров при решении задачи идентификации антропогенных объектов с разной степенью вариативности контуров (границ) на аэрокосмических изображениях. Используются такие нейросетевые модели как U-Net, PSPNet, DeepLabv3+, SegFormer, Twins-PCPVT, ConvNeXt, YOLOv7, YOLOv8, YOLOv9 и YOLOv11.

Исследование включает в себя аннотирование данных космической съемки для определения степени вариативности границ объектов, проведение экспериментов с нейронными сетями и разработку алгоритма сравнения результативности сетей, относящихся к разным типам сегментации, между собой.

Результаты показывают, что нейросетевые модели семантической сегментации более эффективны при обнаружении объектов большой площади, границы которых имеют выраженную вариативность, в то время как модели сегментации экземпляров позволяют получить высокие показатели качества распознавания при работе с объектами, имеющими минимальную вариативность границ.

Ключевые слова: семантическая сегментация, сегментация экземпляров, аэрокосмические изображения, глубокое обучение, метрики, датасет

Введение

Современные высокие темпы развития систем дистанционного зондирования Земли обуславливают применение для интерпретации данных космической съемки методов машинного обучения, в том числе основанных на нейросетевых технологиях. Одним из основных факторов сдерживания успешного развития и применения такого инструментария в программных продуктах, предназначенных для дешифрирования материалов съемки, является сложность прогнозирования качества обучения нейронных сетей (НС). Как правило, такой прогноз осуществляется на основе ранее полученных экспериментальных данных. При этом, важным является вопрос о выборе наиболее подходящего типа и архитектуры НС для решения конкретной задачи. Распознавание и определение границ объектов на снимках традиционно производится с помощью двух типов НС: семантической сегментации и сегментации экземпляров. Выбор наиболее результативного подхода требует от исследователей глубокого понимания специфики задачи и особенностей работы различных НС. Более того, имеются сложности с интерпретацией получаемых результатов, поскольку метрики оценки качества обучения у разных типов НС также отличаются. Поэтому исследования в данном вопросе имеют важное практическое и научное значение. На практике анализ результатов применения различных типов нейросетевых моделей в интересах распознавания антропогенных объектов позволяет в дальнейшем делать более обоснованный выбор архитектуры НС при решении аналогичных задач. Научная составляющая проделанной работы связана с разработкой новой технологии сравнения результативности НС различных типов.

1. Постановка задачи

Задача обнаружения и определения контура (границы) объекта интереса на снимке может быть решена двумя методами дешифрирования, основанными на нейросетевых технологиях, а именно:

- 1) с помощью семантической сегментации (semantic segmentation) – назначение каждому пикселю изображения класса [1];
- 2) с помощью сегментации объекта (instance segmentation) – выявление контуров, каждый из которых соответствует определенному объекту (при этом объекты разделяются на отдельные экземпляры, даже если относятся к одному классу) [2].

Метод семантической сегментации получил свое распространение в таких областях, как медицина, автономное вождение и экологический мониторинг, в то время как сегментация экземпляров находит более широкое применение в сферах робототехники, обработки изображений и видеоданных, а также в торговле и системах безопасности. Однако, применительно к интерпретации данных космической съемки, неизвестно какой из двух рассматриваемых методов окажется наиболее результативным. До сих пор отсутствуют исследования посвященные сравнительному анализу эффективности применения семантической сегментации и сегментации экземпляров при решении задач дешифрирования материалов космической съемки. При этом, интуитивно понятно, что выбор метода сегментации зависит в том числе и от характеристик объектов интереса, одной из которых является вариативность его границ. Анализ работы НС разных типов при распознавании антропогенных объектов с учетом степени вариативности их границ позволит частично ответить на этот вопрос.

Важно отметить, что в проведенном исследовании не ставилась цель поиска оптимальных гиперпараметров для какой-либо нейросетевой модели или получения максимальных значений метрик. Усилия были сосредоточены на том, чтобы понять какие из множества предложенных подходов глубокого обучения «из коробки» (с доступным развёртыванием), со стандартными значениями гиперпараметров, будут максимально результативны при распознавании конкретного класса антропогенных объектов.

Исходные данные

В качестве исходных данных для исследования использовались материалы цветной цифровой космической съемки сверхвысокого и высокого разрешения (от 0.4 м до 2.2 м), в т.ч. приведенные к стандартному виду с помощью панхроматического слияния (pansharpening). Ортотрансформирование данных не производилось, поскольку такое условие, как правило,

приводит к значительному сокращению объема обучающих данных. Всего было аннотировано 2076 цветосинтезированных изображений размером 1280² пикселей.

В качестве объектов исследования выбрано 6 классов: градирни, резервуары, распределительные устройства (РУ), блочные насосные станции (БНС), парковочные зоны (ПЗ) и регулированные сбросы вод (PCB). В таблице 1 приведены основные характеристики обучающего набора данных. В среднем на изображение приходится 10.5 экземпляров объектов, что соответствует плотности эталонных датасетов Cityscapes и ADE20K.

Табл. 1. Характеристики обучающего набора данных

Имя класса/метки	Количество меток	Частота появления на 100 снимков	Общая площадь меток, %	Средняя площадь объектов, пиксели	СКО площадей объектов, пиксели	Процент совпадения с повторной разметкой	Степень вариативности
Градирни	603	15.4	2.8	255.7	73.0	94.6	Минимальная
Резервуары	13024	70.8	59.8	28.3	11.8	80.8	Минимальная
РУ	1924	64.5	8.8	382.1	249.2	83.4	Средняя
БНС	1217	40.7	5.6	136.3	65.8	73.3	Средняя
ПЗ	4216	60.8	19.3	189.1	121.7	76.5	Средняя
PCB	815	22.7	3.7	137.2	121.3	48.1	Выраженная

2. Показатель вариативности границ

Степень вариативности границ объектов различных классов определялась с помощью эксперимента, аналогичному тому, который был описан в работе [3]. Суть эксперимента заключалась в том, что один и тот же человек дважды, с интервалом в несколько месяцев, осуществлял разметку одного и того же набора изображений, после чего два варианта разметки сравнивались между собой. Анализ данных, приведенных в [3], позволяет сделать вывод о том, что во многом величина ошибки (несовпадения контуров при первичной и повторной разметке) зависит от среднего размера объектов класса и от того насколько вариативен их контур.

В эксперименте, проведенном авторами статьи, было переразмечено 10% всех контуров каждого класса присутствующих на 207 изображениях. Временной интервал между оригинальной и повторной разметкой составил около 16 месяцев. Итоговые значения совпадения площадей для классов представлены в столбце «процент совпадения с повторной разметкой» таблицы 1.

Анализируя результаты, можно сделать следующие выводы.

1) Ошибки для градирен незначительны и связаны с неточным определением контура при их активном состоянии: пар затрудняет локализацию контуров.

2) Ошибки разметки БНС можно разделить на две основные группы: пропущенные специфические типы насосных станций и включение/исключение из разметки козлового крана, когда он удален на некотором расстоянии от основных сооружений БНС.

3) Значение 20% в несовпадении площадей резервуаров объясняется их малым размером (см. рисунок 1) и пропусками на изображениях с низким разрешением.

4) Основная ошибка при определении границ РУ связана с разными подходами разметки. В одном случае строился один общий контур для всего объекта, во другом – выделялись несколько локализованных контуров вместо одного общего. При этом, для РУ процент совпадения двух разметок (см. табл. 1) не является показательным с точки зрения определения степени вариативности границ, т.к. сами эти объекты имеют значительную площадь.

5) При разметке ПЗ характерны все перечисленные ранее типы ошибок: например, на рис. 1 вверху присутствует пропуск «парковочной зоны закрытого типа».

6) РСВ характеризуются самым низким показателем площади совпадения, а также самым высоким относительным СКО, как видно из примера на рис. 1, граница таких объектов очень условна и зависит не только от показателей качества изображения (линейное разрешение на местности, степень освещенности, облачность), но и от индивидуальных особенностей разметчика.

Совокупность показателей «средняя площадь объектов», «СКО площадей объектов» и «процент совпадения с повторной разметкой», полученных в ходе проведенного эксперимента, позволяет условно разделить все рассматриваемые классы объектов по степени вариативности их контуров (границ) на 3 группы. Объекты с минимальной степенью вариативности границ имеют четкие границы и устойчивые простые формы. Объекты со средней степенью вариативности границ обладают сложными, неустойчивыми формами, а результат выделения их контуров зависит от индивидуальных особенностей восприятия сцены оператором. Объекты с выраженной степенью вариативности характеризуются изменением своей формы с течением времени и нечеткостью своих границ: наличием буферной зоны между объектом и фоном в пределах которой затруднительно однозначно определить принадлежность каждого конкретного пикселя к объекту или фону.



■ Резервуары, ■ БНС, ■ ПЗ, ■ PCB, ■ РУ

Рис. 1. Сравнение оригинальной (темный контур) и повторной (светлый контур) разметок (полупрозрачной заливкой отмечена площадь пересечения)

3. Нейросетевые модели и результаты обучения

В данной работе исследовались наиболее значимые и популярные НС, обеспечивающие сегментацию изображения. Для большей конкретики в статье приводятся названия выбранных моделей, дата их выпуска и, в случае моделей семантической сегментации, их базовая модель (backbone).

В качестве моделей семантической сегментации использовались следующие: U-Net [4], backbone «UNet-S5-D16» (2015.05), PSPNet [5], backbone «R-101-D8» (2017.04), DeepLabv3+ [6], backbone «R-101-D8» (2018.08), SegFormer [7], backbone «MIT-B4» (2021.05), Twins-PCPVT [8], размер «pcpvt-l» (2021.09), ConvNeXt [9], размер модели «large» (2022.03).

НС сегментации экземпляров были представлены следующими моделями: YOLOv7 размер «yolov7x-seg», [10] (2022.07), YOLOv8 [11], размер «yolov8x-seg» (2023.01), YOLOv9 [12], размер «yolov9e-seg», (2024.02), YOLOv11 [11], размер «yolo11x-seg» (2024.09).

Популярная модель SAM [13] не рассматривалась, поскольку, как справедливо отмечено в [14], она спроектирована для сегментации экземпляров, но при этом не формирует ассоциативной связи с создаваемыми контурами, что является ключевым условием для рассматриваемой задачи распознавания антропогенных объектов.

Для исследования НС семантической сегментации была использована библиотека MMSegmentation [15], для НС сегментации экземпляров YOLOv8 и YOLOv11 – библиотека Ultralytics [16] и отдельные реализации для моделей YOLOv7 [17] и YOLOv9 [16]. Интерес исследования свойств новых версий сетей YOLO возникает потому, что традиционно Ultralytics не выпускают сопровождающих статей при их релизе, но научное сообщество с энтузиазмом делает это за них [18-20].

Включение в список сравнительно старых НС PSPNet и U-Net связано с тем, что их архитектура и решения послужили фундаментом для развития современных моделей, таким образом, это позволит оценить, насколько далеко продвинулись новые сетевые модели и насколько результативен «базовый» подход семантической сегментации.

Многочисленные исследования на тему аугментации [21-23], свидетельствуют о том, что применение этой процедуры позволяет повысить качество обучения НС. В данной работе настройка аугментации производилась таким образом, чтобы она обеспечивала равные условия обучения для всех исследуемых моделей НС, а именно, использовались яркостные (изменение яркости, контраста, насыщенности, оттенка) и геометрические (случайное изменение размера, случайное кадрирование, случайное отражение слева направо и сверху вниз) параметры аугментации.

Размер пакета нормализации (batch size) для всех методов использовался равным 4. В исследовании [3] сообщается, что этот показатель имеет решающее значение при обучении НС, однако имеющиеся технические ограничения не позволили исследовать этот вопрос.

Предварительные пробные запуски обучения на базе MMSegmentation позволили сделать вывод, что наиболее качественный результат получается при сохранении исходного разрешения. При обучении НС был использован планировщик (schedule) 240000, который обеспечивал выход на плато для всех анализируемых моделей семантической сегментации. Для моделей сегментации экземпляров количество эпох было установлено равным 240.

Валидация моделей производилась на тестовой выборке из 200 изображений.

Результаты обучения

Стандартные показатели качества обучения НС, рассчитанные по тестовому набору данных приведены в таблицах: 2 – для сетей сегментации экземпляров, 3 – для сетей семантической сегментации. При этом, метрики, представленные в таблице 2 рассчитывались для сегментов (не для ограничивающих рамок типа box) распознанных объектов.

Табл. 2. Показатели качества для обученных моделей сегментации экземпляров

Модель	Градири		Резервуары		РУ		БНС		ПЗ		PCB		Среднее по всем классам	
	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95
YOLOv7	1.00	0.84	0.90	0.60	0.67	0.41	0.65	0.33	0.49	0.29	0.32	0.10	0.67	0.43
YOLOv8	0.99	0.84	0.91	0.59	0.77	0.47	0.83	0.47	0.60	0.36	0.38	0.12	0.75	0.48
YOLOv9	1.00	0.84	0.90	0.55	0.79	0.47	0.84	0.46	0.58	0.36	0.27	0.10	0.73	0.46
YOLOv11	1.00	0.85	0.90	0.52	0.82	0.48	0.82	0.43	0.56	0.33	0.20	0.07	0.71	0.45

Табл. 3. Показатель точности (ассигасу) для обученных моделей семантической сегментации

Модель	Градири	Резервуары	РУ	БНС	ПЗ	PCB	Среднее по всем классам
U-Net	67.3	75.1	80.2	23.2	66.6	19.0	61.5
PSPNet	90.6	74.9	88.9	38.8	65.3	23.9	68.8
DeepLabv3+	90.0	71.8	88.9	54.3	66.3	35.2	72.2
SegFormer	92.0	56.1	85.3	29.7	62.9	25.1	64.3
Twins-PCPVT	94.0	80.4	88.3	52.0	70.5	47.0	75.8
ConvNeXt	93.1	77.9	92.5	57.0	80.5	48.1	78.3

Примеры результатов предсказаний на тестовых изображениях обученных моделей НС приведены на рисунках 2-4.

Анализ значений показателей качества обучения НС и визуальный анализ результатов распознавания позволил сделать следующие выводы.

- 1) Среднее значение метрики ассигасу по всем классам для моделей семантической сегментации свидетельствует об устойчивой тенденции роста качества моделей с развитием методов глубокого обучения. Сравнительно низкая величина этого показателя для SegFormer может объясняться некорректным выбором базовой модели.
- 2) Отмечается корреляция между показателями качества обучения НС и степенью вариативности границ распознаваемых объектов: объекты PCB имеют самые низкие показатели для обоих типов НС. Градири, напротив, несмотря на меньшее по сравнению с PCB количество меток, распознаются более уверенно.
- 3) На показатели качества обучения НС также влияют геометрические размеры распознаваемых объектов, что подтверждается работой [3].

4) На примере сегментации РУ (рисунок 2), видно, что с задачей распознавания объекта, в случае если он находится в тени, наиболее результативно справляются НС сегментации экземпляров. Из сетей семантической сегментации удовлетворительные результаты показала только НС ConvNeXt несмотря на то, что архитектура НС Twins относится к более современным сетям – трансформерам.

4. Способ сравнения нейросетей различных типов

Сравнение результативности распознавания объектов сетями сегментации экземпляров и семантической сегментации по соответствующим метрикам невозможно, так как эти метрики имеют принципиально разные алгоритмы расчета. Расчет mAP50 и mAP50-95 основывается на количестве правильно и ошибочно распознанных объектов, а также учитывается степень уверенности сети отнесения распознанного объекта к данному классу (confidence). НС семантической сегментации не разделяет результаты распознавания на отдельные объекты, поэтому при применении данного метода получается другой набор выходных данных, не позволяющий рассчитать метрику AP. Вместе с тем, поиск метрик наиболее релевантно оценивающих качество обучения НС разных типов продолжается до сих пор. На сегодняшний день этот процесс нельзя назвать завершенным: предлагаются различные подходы расчета показателей результативности распознавания как для семантической сегментации [24], так и для сегментации экземпляров [25]. Это обстоятельство позволяет предложить свой, уникальный способ оценки работы НС разных типов, который обеспечивал бы их сравнение.

Для сравнения работы НС разных типов между собой наиболее подходящей является стандартная метрика «пересечения на объединение» (IoU), которая может быть рассчитана как при решении задачи сегментации экземпляров, так и при семантической сегментации и позволяет оценить степень совпадения между предсказанными и эталонными границами. Однако, алгоритмы расчета этой метрики у сетей разных типов также отличаются.

В задаче сегментации экземпляров метрика IoU представляет собой степень перекрытия эталонного и предсказанного контуров для каждого объекта в отдельности и может сильно изменяться в зависимости от установленного порога confidence и значения подавления не-максимумов (non-maximum suppression) – параметра, который отвечает за частоту построения перекрывающихся контуров распознаваемых объектов. При этом могут пересекаться границы объектов, принадлежащих как разным классам, так и к одному классу.

В задаче семантической сегментации такие настройки НС, как confidence и non-maximum suppression не имеют значения. Кроме того, отсутствует эффект перекрытия контуров от разных объектов.

Таким образом, при сравнении результативности НС разных типов по метрике IoU существуют две проблемы:

- в результатах распознавания с помощью НС сегментации экземпляров необходимо учесть эффект пересечения контуров объектов интереса, который отсутствует у сетей семантической сегментации,
- при обработке материалов съемки с помощью НС сегментации экземпляров необходимо обосновать выбор параметров confidence и non-maximum suppression, которые отсутствуют у НС семантической сегментации.

Для решения первой проблемы, перед расчетом метрики IoU для НС сегментации экземпляров, результаты распознавания модифицируются следующим образом.

- 1) Все экземпляры объектов объединяются по классам.
- 2) При перекрытии контуров объектов, относящихся к одному классу, выполняется слияние этих контуров.

Метрика IoU таким образом может быть рассчитана аналогично моделям семантической сегментации. При таком подходе теряется ключевая идея работы алгоритма сегментации экземпляров – разделение контуров объектов внутри класса, но, с другой стороны, это позволяет в численном эквиваленте сравнить результативность НС различных типов.

Для решения второй проблемы были проведены исследования изменения метрики IoU, рассчитанной по модифицированным результатам распознавания в зависимости от параметров confidence и non-maximum suppression. Полученные значения для НС YOLOv11 приведены в таблице 4. Закономерность изменения IoU для других НС семейства YOLO имеет схожий характер.

При разработке способа сравнения НС, вариант пересечения контуров объектов, относящихся к разным классам, рассматривался, однако в нашем датасете подобные примеры не встречались и соответствующих исследований не проводилось. Тем не менее указанные случаи безусловно влияют на конечные значения IoU и исследование этого вопроса представляет практический интерес.

Анализ приведенных в таблице данных позволяет определить наиболее оптимальные значения confidence и non-maximum suppression равными 0,1 и 0,5 соответственно. Объясняется это следующим образом.

Для confidence, в интервале от 0,1 до 0,9 с ростом значений этого параметра наблюдается снижение IoU. Такой эффект объясняется увеличением числа ложноотрицательных сегментов. В это же время, уменьшение confidence от 0,1 до 0,05 также приводит к снижению IoU, однако в данном случае это происходит за счет роста ложно положительных объектов. Таким образом, максимум IoU регистрируется при confidence равном 0,1.

Табл. 4. Изменение значений IoU в зависимости от параметров confidence и non-maximum suppression, для модели YOLOv11

		<i>non-maximum suppression</i>									
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
<i>confidence</i>	0.05	59.5	62.3	62.6	62.8	62.8	63.0	62.9	62.9	62.9	63.0
	0.1	59.9	62.7	62.9	63.2	63.4	63.5	63.4	63.5	63.5	63.6
	0.2	59.1	62.0	62.1	62.3	62.5	62.7	62.6	62.7	62.7	62.8
	0.3	58.1	61.0	61.2	61.3	61.5	61.7	61.7	61.6	61.7	61.7
	0.4	56.7	59.5	59.7	59.9	59.9	60.1	60.2	60.2	60.2	60.2
	0.5	54.7	57.6	57.8	57.9	58.0	58.1	58.1	58.1	58.1	58.1
	0.6	52.4	55.2	55.4	55.5	55.6	55.7	55.7	55.7	55.7	55.7
	0.7	48.3	50.9	51.1	51.2	51.2	51.3	51.3	51.3	51.3	51.3
	0.8	41.2	43.2	43.2	43.3	43.3	43.4	43.4	43.4	43.4	43.4
	0.9	15.7	16.2	16.2	16.2	16.2	16.2	16.2	16.2	16.2	16.2

При выборе значения параметра non-maximum suppression очевидно, что приравнивание его 0 (отсутствие пересекающихся контуров) приведет к ощутимой потере качества распознавания, поскольку при пересечении границ объектов алгоритм выбирает только самый значимый контур, остальные с которыми происходит пересечение удаляются из результатов распознавания. Стабилизация результатов происходит при значениях, больше 0,4.

5. Результаты исследования

С учетом обоснованного выбора значений параметров confidence и non-maximum suppression и преобразования результатов распознавания для НС сегментации экземпляров, была получена возможность сравнить по метрике IoU результативность НС различных типов при распознавании антропогенных объектов. Полученные значения метрики представлены в таблице 5.

Анализ приведенных в таблице 5 данных позволил сделать следующие выводы.

При распознавании антропогенных объектов наиболее результативной является НС ConvNeXt. Она имеет наиболее высокий усредненный показатель IoU среди всех моделей. Кроме того, эта сеть показывает максимальные значения IoU по сравнению с другими исследуемыми моделями семантической сегментации практически для всех классов объектов, за исключением резервуаров. Однако и в этом случае разница между максимальным значением равным 73,5 которое достигается с помощью НС Twins-PCPVT, и значением 69,4 при использовании ConvNeXt допустима, применительно к рассматриваемой задаче.

Антропогенные объекты, характеризующиеся средней и выраженной степенью вариативности границ (за исключением БНС), наиболее достоверно распознаются с помощью НС семантической сегментации. Это подтверждается как максимальными значениями IoU, так и сравнением средних значений этого показателя, рассчитанных по всем НС, принадлежащих одному типу, для каждого класса объектов. Сложность распознавания БНС сетями

семантической сегментации может быть связана с нехваткой обучающих данных: объекты этого класса характеризуются высокой вариативностью своих конструкций, при этом по количеству меток они находятся на 3 месте в наборе обучающих данных, ниже расположены только РСВ, имеющие на снимках устойчивые яркостные и текстурные характеристики, и градирни – сравнительно большие объекты имеющие устойчивые геометрические формы и размеры.

Табл. 5. Показатели качества обучения для апробированных моделей по метрике IoU

Модель	градирня	резервуар	РУ	БНС	ПЗ	РСВ	Среднее по всем классам
U-Net	60,1	67,4	74,9	22,5	61,2	17,6	50,6
PSPNet	69,9	69,3	82,2	36,6	61,5	23,3	57,1
DeepLabv3+	87,1	67,2	81,4	48,1	60,4	29,4	62,3
SegFormer	78,0	50,3	77,9	26,6	57,8	20,7	51,9
Twins-PCPVT	72,3	73,5	80,7	44,6	64,4	40,1	62,6
ConvNeXt	87,6	69,4	84,4	52,5	71,7	40,9	67,8
YOLOv7	90,8	68,6	62,4	43,8	54,2	23,6	57,2
YOLOv8	88,7	66,0	70,2	58,5	63,2	37,3	64,0
YOLOv9	89,4	65,6	70,0	56,4	62,3	27,5	61,9
YOLOv11	89,2	65,6	72,3	60,1	62,3	30,8	63,4

Крупные объекты с минимальной вариативностью (градирни) лучше распознаются с помощью моделей семейства YOLO. Для всех НС этого типа значение показателя IoU не опускается ниже 88, а максимальный результат, равный почти 91 показывает сеть YOLOv7. При этом результаты НС семантической сегментации (как и разница между их выпуском) варьируются в достаточно широких пределах: от 60 до 87.

Резервуары, являющиеся небольшими объектами с минимальной степенью вариативности границ, в целом с соразмеряемой достоверностью распознаются как НС семантической сегментации, так и сегментации экземпляров. Чуть более низкие показатели IoU у последних объясняются перекрытием областей расположения у близко размещенных друг к другу объектов.

Низкие показатели IoU при распознавании РСВ имеют несколько причин: самое высокое отношение СКО/средней площади, низкая частота появления на снимках и выраженная вариативность контуров (т.е. самое малое значение совпадения контуров при повторной разметке).

Заключение

По результатам исследований можно сделать вывод, что степень вариативности границ объектов, наряду с другими характеристиками класса, является важным, но не до конца проработанным и изученным показателем. Учет этого параметра может помочь при решении

различных задач компьютерного зрения уже на этапе подготовки данных. Так, например, для объектов с минимальной степенью вариативности границ достаточно сравнительно небольшого количества примеров в обучающем наборе данных для достижения высоких показателей результативности распознавания. Основное внимание при этом следует сосредотачивать на объектах с выраженной степенью вариативности границ.

До сих пор не было проведено исследований по сравнению методов сегментации экземпляров и семантической сегментации. С рядом оговоренных условий, представленные в статье материалы позволяют решить эту задачу.

Вместе с тем, в дальнейшем хотелось бы исследовать применимость моделей НС к более широкому классу объектов, с учетом степени вариативности их границ. Также необходимо оценить влияние гиперпараметров для улучшения обучения моделей. Кроме того, большой интерес представляет анализ методов паноптической сегментации, которые мы на данном этапе обошли стороной.

Благодарности

Авторы выражают особую благодарность Роману Хабарову за вдохновение и идеи, без которых бы не было данного исследования

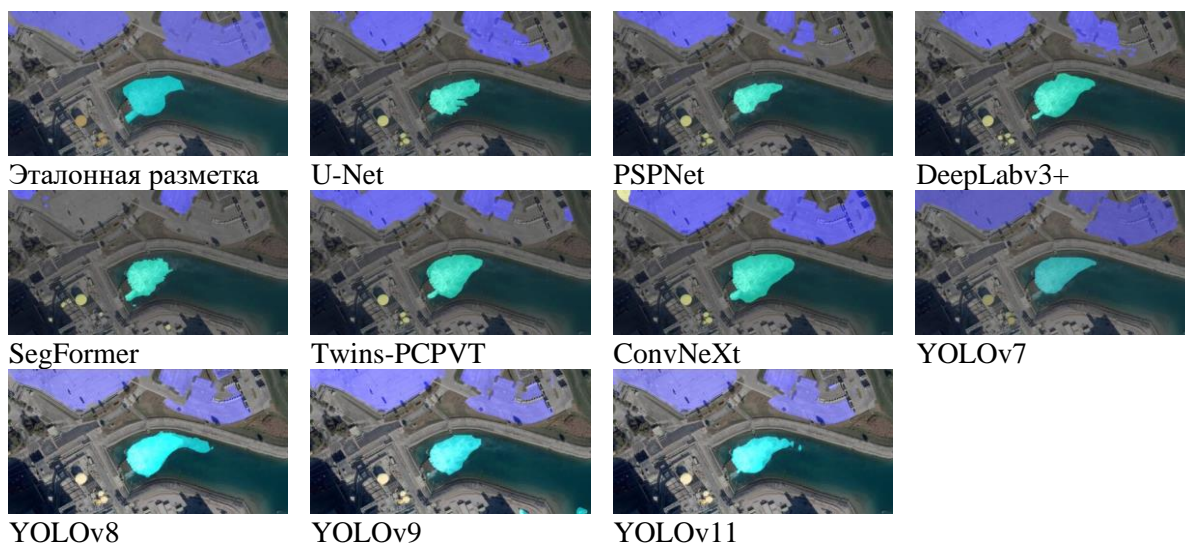


Рисунок 2. Результаты обработки моделей



Рис. 3. Результаты обработки моделей

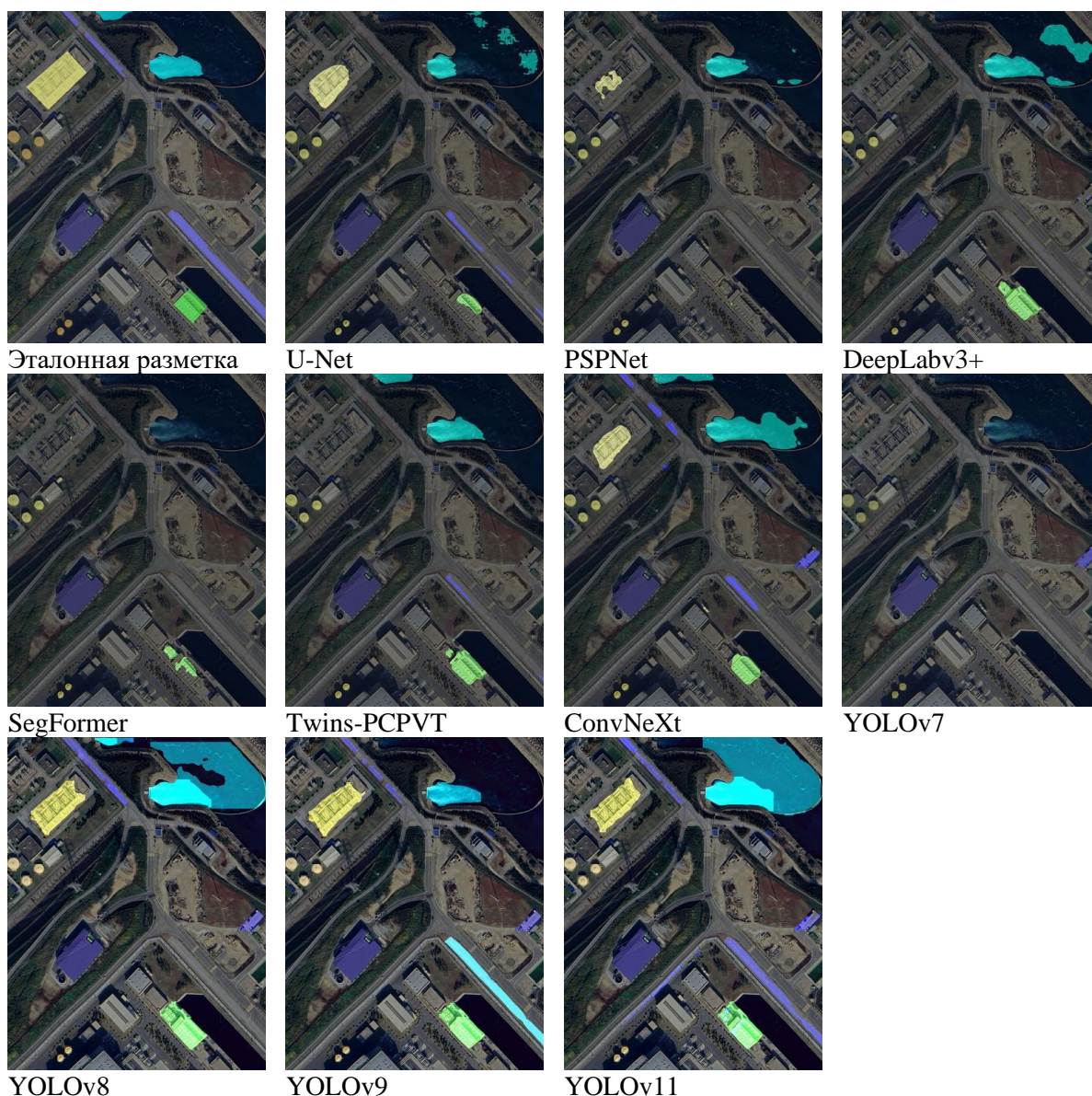


Рисунок 4. Результаты обработки моделей

Литература:

1. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation // CVPR, IEEE Computer Society, 2015, P. 3431-3440.
2. He K., Gkioxari G., Dollar P. and Girshick R., Mask R-CNN // IEEE International Conference on Computer Vision, Venice, Italy, 2017. P. 2980-2988.
3. Zhao H., Puig X., Xiao T., Fidler S., Barriuso A., Torralba A. Semantic Understanding of Scenes through the ADE20K Dataset. 2016 // Electronic resource. URL: <http://arXiv.org> (accessed 11.07.2024).
4. Olaf Ronneberger, Philipp Fischer, & Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 // Electronic resource. URL: <http://arXiv.org> (accessed 11.07.2024).
5. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, Pyramid scene parsing network // IEEE Computer Vision and Pattern Recognition Conference, Honolulu, Hawaii, 2017. P. 6230-6239.
6. Chen L.-C. [et al.] Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. 2018. // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).
7. Xie E. [et al.]. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. 2021. // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).

8. Chu X., Tian Z., Wang Y., Zhang B., Ren H., Wei X., Xia H., Shen Ch., Twins: Revisiting the Design of Spatial Attention in Vision Transformers // *Advances in Neural Information Processing Systems* 34, 2021.
9. Z. Liu, H. Mao, C. -Y. Wu, C. Feichtenhofer, T. Darrell and S. Xie. A ConvNet for the 2020s // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022. P.11966-11976.
10. Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors // IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023. P. 7464-7475.
11. Jocher, G., Qiu, J., & Chaurasia, A. Ultralytics YOLO [Computer software] // Electronic resource. URL: <https://github.com/ultralytics/ultralytics> (accessed 02.09.2024).
12. Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. 2024 // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).
13. Kirillov A. [et al.]. Segment Anything // 2023 IEEE/CVF International Conference on Computer Vision, Paris, France, 2023. P. 3992-4003.
14. Jieli Zhang, Zhongliang Zhou, Gengchen Mai, Lan Mu, Mengxuan Hu, Sheng Li, Text2Seg: Remote Sensing Image Semantic Segmentation via Text-Guided Visual Foundation Models. 2023. // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).
15. K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin. MMDetection: Open MMLab detection toolbox and benchmark. 2019 // Electronic resource. URL: <http://arXiv.org> (accessed 12.04.2023).
16. Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao, Implementation of paper - YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information // Electronic resource. URL: <https://github.com/WongKinYiu/yolov9> (accessed 24.12.2024).
17. Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao, Implementation of paper - YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors // Electronic resource. URL: <https://github.com/WongKinYiu/yolov7> (accessed 04.08.2024).
18. Mupparaju Sohan, Thotakura Sai Ram, Venkata Rami Reddy Ch, A Review on YOLOv8 and its Advancements // *Data Intelligence and Cognitive Informatics*. 2024. DOI:10.1007/978-981-99-7962-2_39. P. 529-545.
19. Juan Terven, Diana Cordova-Esparza, A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS // *Machine Learning and Knowledge Extraction*. 2023. V. 5(4) DOI:10.3390/make5040083. P. 1680-1716.
20. Ankan Ghosh, YOLO11: Faster Than You Can Imagine! // Electronic resource. URL: <https://learnopencv.com/yolo11> (accessed 15.10.2024).
21. Shorten C., Khoshgoftaar T. M. A survey on Image Data Augmentation for Deep Learning // *Journal of Big Data*. 2019. No.1 (6). P. 60.
22. Zoph B. [et al.]. Learning Data Augmentation Strategies for Object Detection // Electronic resource. URL: <http://arXiv.org> (accessed 29.04.2024).
23. Ghaffar M. A. A. [et al.]. Data augmentation approaches for satellite image super-resolution // *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2019. Vol. IV-2/W7. P. 47–54.
24. Angelo Monteux, Metrics for semantic segmentation // Electronic resource. URL: <https://ilmonteux.github.io/2019/05/10/segmentation-metrics.html> (accessed 14.11.2024).
25. Chen L., Wu Y., Stegmaier J., Merhof D., SortedAP: Rethinking evaluation metrics for instance segmentation // 2023 IEEE/CVF International Conference on Computer Vision Workshop, DOI:10.1109/ICCVW60793.2023.00424, 2023. P. 3925-3931.

Features of Neural Network Application in Recognizing Anthropogenic Objects with Variable Contours

A.G. Saidov., D.V. Zhukov¹

¹Mozhaisky Military Space Academy, Russia, St. Petersburg

Abstract. The study analyzes the effectiveness of semantic segmentation and instance segmentation methods for identifying anthropogenic objects with varying degrees of boundary variability in aerospace imagery. Neural network models such as U-Net, PSPNet, DeepLabv3+, SegFormer, Twins-PCPVT, ConvNeXt, YOLOv7, YOLOv8, YOLOv9, and YOLOv11 are utilized.

The research involves annotating remote sensing data to determine the degree of boundary variability, conducting experiments with neural networks, and developing an algorithm to compare the performance of networks belonging to different segmentation types.

The results demonstrate that semantic segmentation models are more effective for detecting large-area objects with pronounced boundary variability, while instance segmentation models achieve high recognition accuracy for objects with minimal boundary variability.

Keywords: semantic segmentation, instance segmentation, aerospace imagery, deep learning, metrics, dataset

Table 1. Characteristics of the training dataset

Table 2. Metrics for trained instance segmentation models

Table 3. Accuracy scores for trained semantic segmentation models

Table 4. Changes in IoU values depending on confidence and non-maximum suppression parameters for the YOLOv11 model

Table 5. Training quality metrics for evaluated models (IoU metric)

Fig. 1. Comparison of original (dark outline) and relabeled (light outline) annotations (semi-transparent fill indicates the intersection area)

Figures 2–4. Model inference results

References:

1. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation // CVPR, IEEE Computer Society, 2015, P. 3431-3440.
2. He K., Gkioxari G., Dollar P. and Girshick R., Mask R-CNN // IEEE International Conference on Computer Vision, Venice, Italy, 2017. P. 2980-2988.
3. Zhao H., Puig X., Xiao T., Fidler S., Barriuso A., Torralba A. Semantic Understanding of Scenes through the ADE20K Dataset. 2016 // Electronic resource. URL: <http://arXiv.org> (accessed 11.07.2024).
4. Olaf Ronneberger, Philipp Fischer, & Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 // Electronic resource. URL: <http://arXiv.org> (accessed 11.07.2024).
5. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, Pyramid scene parsing network // IEEE Computer Vision and Pattern Recognition Conference, Honolulu, Hawaii, 2017. P. 6230-6239.
6. Chen L.-C. [et al.] Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. 2018. // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).
7. Xie E. [et al.]. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. 2021. // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).
8. Chu X., Tian Z., Wang Y., Zhang B., Ren H., Wei X., Xia H., Shen Ch., Twins: Revisiting the Design of Spatial Attention in Vision Transformers // Advances in Neural Information Processing Systems 34, 2021.
9. Z. Liu, H. Mao, C. -Y. Wu, C. Feichtenhofer, T. Darrell and S. Xie. A ConvNet for the 2020s // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022. P.11966-11976.
10. Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors // IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023. P. 7464-7475.
11. Jocher, G., Qiu, J., & Chaurasia, A. Ultralytics YOLO [Computer software] // Electronic resource. URL: <https://github.com/ultralytics/ultralytics> (accessed 02.09.2024).
12. Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. 2024 // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).
13. Kirillov A. [et al.]. Segment Anything // 2023 IEEE/CVF International Conference on Computer Vision, Paris, France, 2023. P. 3992-4003.

14. Jielu Zhang, Zhongliang Zhou, Gengchen Mai, Lan Mu, Mengxuan Hu, Sheng Li, Text2Seg: Remote Sensing Image Semantic Segmentation via Text-Guided Visual Foundation Models. 2023. // Electronic resource. URL: <http://arXiv.org> (accessed 14.07.2024).
15. K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin. MMDetection: Open MMLab detection toolbox and benchmark. 2019 // Electronic resource. URL: <http://arXiv.org> (accessed 12.04.2023).
16. Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao, Implementation of paper - YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information // Electronic resource. URL: <https://github.com/WongKinYiu/yolov9> (accessed 24.12.2024).
17. Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao, Implementation of paper - YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors // Electronic resource. URL: <https://github.com/WongKinYiu/yolov7> (accessed 04.08.2024).
18. Mupparaju Sohan, Thotakura Sai Ram, Venkata Rami Reddy Ch, A Review on YOLOv8 and its Advancements // Data Intelligence and Cognitive Informatics. 2024. DOI:10.1007/978-981-99-7962-2_39. P. 529-545.
19. Juan Terven, Diana Cordova-Esparza, A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS // Machine Learning and Knowledge Extraction. 2023. V. 5(4) DOI:10.3390/make5040083. P. 1680-1716.
20. Ankan Ghosh, YOLO11: Faster Than You Can Imagine! // Electronic resource. URL: <https://learnopencv.com/yolo11> (accessed 15.10.2024).
21. Shorten C., Khoshgoftaar T. M. A survey on Image Data Augmentation for Deep Learning // Journal of Big Data. 2019. No.1 (6). P. 60.
22. Zoph B. [et al.]. Learning Data Augmentation Strategies for Object Detection // Electronic resource. URL: <http://arXiv.org> (accessed 29.04.2024).
23. Ghaffar M. A. A. [et al.]. Data augmentation approaches for satellite image super-resolution // ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019. Vol. IV-2/W7. P. 47–54.
24. Angelo Monteux, Metrics for semantic segmentation // Electronic resource. URL: <https://ilmonteux.github.io/2019/05/10/segmentation-metrics.html> (accessed 14.11.2024).
25. Chen L., Wu Y., Stegmaier J., Merhof D., SortedAP: Rethinking evaluation metrics for instance segmentation // 2023 IEEE/CVF International Conference on Computer Vision Workshop, DOI:10.1109/ICCVW60793.2023.00424, 2023. P. 3925-3931.

About author(s)

Saidov Alisher G. Scientific researcher. Research areas: computer vision, artificial intelligence, data processing, geographic information system. E-mail: mysnow@mail.ru.

Zhukov Denis V. Candidate of technical sciences, senior scientific researcher. Senior scientific researcher, Mozhaisky Military Space Academy. Research areas: computer vision, satellite data processing, artificial intelligence, information mining. E-mail: spb.pilligrim@gmail.com.