



Особенности применения нейронных сетей при распознавании антропогенных объектов с вариативными контурами

А.Г. Саидов¹, Д.В. Жуков¹✉

¹ Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, Россия
✉ spb_pilligrim83@mail.ru

ЦИТИРОВАНИЕ Саидов А.Г., Жуков Д.В. Особенности применения нейронных сетей при распознавании антропогенных объектов с вариативными контурами // Известия вузов «Геодезия и аэрофотосъемка». 2025. Т. 69, № 4. С. 83–95. DOI:10.30533/GiA-2025-030.

КЛЮЧЕВЫЕ СЛОВА семантическая сегментация, сегментация экземпляров, аэрокосмические изображения, глубокое обучение, метрики, датасет

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. Авторы выделяют три уровня вариативности контуров объектов и анализируют ее влияние на точность и обобщаемость моделей. Важным аспектом работы является анализ связи между вариативностью контура объекта и эффективностью используемых методов глубокого обучения. 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. Обсуждаются метрики оценки качества сегментации и особенности их интерпретации. 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. Подчеркивается важность учета вариативности контуров объектов при подготовке данных и выборе методов сегментации, а также необходимость дальнейших исследований для улучшения обучения моделей и повышения вероятности выявления объектов.

1 Введение

Высокие темпы современного развития космических систем дистанционного зондирования Земли обуславливают обращение для интерпретации данных, получаемых от этих систем, к методам машинного обучения, в том числе основанным на нейросетевых технологиях. Одним из основных факторов, который сдерживает успешное использование такого инструментария в программных продуктах тематической обработки космических изображений (КИ), является сложность прогнозирования качества обучения нейронных сетей (НС). Как правило, такой прогноз осуществляется на основе ранее полученных экспериментальных данных. При этом на первом этапе разработки модели НС важным является вопрос о выборе наиболее подходящего типа и архитектуры НС. Сложности, возникающие при решении данного вопроса экспериментальным путем, связаны с интерпретацией получаемых результатов, поскольку метрики оценки качества обучения у разных типов НС также различаются. Таким образом, исследования данной проблемы имеют важное практическое и научное значение. На практике анализ результатов применения различных НС в интересах распознавания антропогенных объектов на КИ (одна из основных задач тематической обработки) позволяет в дальнейшем делать более обоснованный выбор типа и архитектуры НС. Научная составляющая связана с разработкой новой технологии сравнения результативности НС различных типов.

Задача распознавания и определения границы объекта интереса на КИ традиционно решается с помощью НС двух типов:

- 1) семантической сегментации (*англ.* semantic segmentation) — назначение каждому пикселю изображения определенного класса [1];
- 2) сегментации экземпляров (*англ.* instance segmentation) — определение границы каждого предварительно распознанного объекта (при этом объекты представляют собой отдельные экземпляры, даже если относятся к одному классу) [2].

Оба типа НС широко используются в различных сферах жизни общества: медицине, автономном вождении, робототехнике и т. д. Однако до сих пор отсутствуют исследования, посвященные сравнительному анализу НС применительно к дешифрированию КИ. При этом выбор типа НС зависит в том числе и от характеристик объектов интереса. Одной из них является вариативность границ, под которой следует понимать не изменение контура объекта с течением времени, а его четкость и однозначность определения на КИ. Влияние этого параметра на результативность различных НС ранее не изучалось.

В данной работе была поставлена задача провести анализ качества обучения НС разных типов и архитектур при распознавании антропогенных объектов с учетом степени вариативности их границ. При этом не ставилась цель получения максимальных значений показателей качества обучения НС. Усилия были сосредоточены на том, чтобы понять, какие из множества доступных на сегодняшний день НС со стандартными значениями гиперпараметров будут наиболее результативно распознавать конкретный класс антропогенных объектов.

2 Материалы и методы

2.1 Набор данных для обучения нейронных сетей

Для решения поставленной задачи был сформирован набор обучающих данных, включающий в себя материалы трехканальной цветной цифровой космической съемки сверхвысокого и высокого разрешения (от 0,4 до 2,2 м), предоставляемые различными открытыми геосервисами, такими как «Яндекс Карты», ESRI World Imagery, Bing Maps и т. д. Всего было аннотировано 2076 КИ размером 12 802 пикселей.

В качестве объектов интереса было выбрано 6 классов: градири, резервуары, распределительные устройства (РУ), блочные насосные станции (БНС), парковочные зоны (ПЗ) и регулированные сбросы вод (РСВ). В табл. 1 приведены основные характеристики обучающего набора данных.

Класс	Кол-во меток	Частота появления на сто снимков	Общая площадь меток, %	Средняя площадь объектов, пиксели	Среднеквадратичное отклонение площадей объектов, пиксели	Совпадение с повторной разметкой, %	Степень вариативности
Градири	603	15,4	2,8	255,7	73,0	94,6	Минимальная
Резервуары	13 024	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
РСВ	815	22,7	3,7	137,2	121,3	48,1	Выраженная

Таблица 1 ⬆
Характеристики обучающего набора данных

Table 1
Characteristics of the training dataset

2.2 Метод определения степени вариативности границ объектов интереса

Для определения количества градаций степени вариативности границ объектов различных классов был проведен эксперимент, аналогичный тому, который 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 полученных в указанной работе результатов позволяет сделать вывод, что во многом величина ошибки (степень несовпадения контуров при первичной и повторной разметке) зависит от среднего размера объектов в классе и от вариативности их границ.

В эксперименте, проведенном авторами статьи, было переразмечено 207 изображений, которые содержали примерно по 10 % от изначального количества объектов каждого класса в сформированном наборе данных, представленном в табл. 1. Итоговые значения совпадения площадей для классов указаны в столбце «Совпадение с повторной разметкой». Примеры первичной и повторной разметки приведены на рис. 1.

Рис. 1 ⬆
Сравнение оригинальной и повторной разметок

Fig. 1
Comparison of original and relabeled annotations



Анализ результатов позволяет сделать следующие выводы.

1. Неточности определения границ для градиен незначительны и связаны с трудностями определения контура при активном состоянии объектов из-за пара.
2. При разметке БНС оператор мог включить в контур или исключить из него козловой кран при удалении последнего от основных сооружений БНС.
3. Значение 20 % в несовпадении площадей резервуаров объясняется пропусками объектов, имеющих малый размер, на КИ с низким разрешением (рис. 1).
4. При определении границ РУ использовались разные подходы. В одном случае строился один общий контур для всего объекта, в другом — выделялись несколько локализованных контуров вместо одного общего. При этом малое значение процента совпадения разметок (табл. 1) объясняется тем, что РУ имеют значительную площадь.
5. Для разметки ПЗ характерны такие типовые ошибки, как пропуск объектов и разные подходы при определении границ.
6. РСВ характеризуются самым низким показателем площади совпадения, а также самым высоким относительным среднеквадратичным отклонением; как видно из примера на рис. 1, граница таких объектов очень условна.

Совокупность таких показателей, как средняя площадь объектов, среднеквадратичное отклонение площадей объектов и процент совпадения с повторной разметкой, полученных в ходе проведенного эксперимента, позволяет условно разделить все рассматриваемые классы объектов по степени вариативности их границ на три группы. К первой группе (минимальная степень вариативности) относятся объекты, имеющие четкие границы и простые формы. Объекты второй группы (средняя степень вариативности) обладают сложными формами, а результат выделения их контуров зависит от особенностей восприятия сцены оператором. Объекты третьей группы (выраженная степень вариативности) характеризуются нечеткостью своих границ: наличием буферной зоны между объектом и фоном, в пределах которой затруднительно однозначно определить принадлежность каждого конкретного пикселя к объекту или фону.

2.3 Используемые архитектуры нейронных сетей и их настройка

В данной работе исследовались наиболее значимые и популярные НС, обеспечивающие сегментацию изображения. В качестве моделей семантической сегментации использовались: U-Net backbone UNet-S5-D16 (2015.05) [4], PSPNet backbone R-101-D8 (2017.04) [5], DeepLabv3+ backbone R-101-D8 (2018.08) [6], SegFormer backbone MIT-B4 (2021.05) [7], Twins-PCPVT (2021.09) [8], ConvNeXt (2022.03) [9]. НС сегментации экземпляров были представлены следующими моделями: YOLOv7 (2022.07) [10], YOLOv8 (2023.01)¹, YOLOv9 (2024.02) [11], YOLOv11 (2024.09) [11].

Интерес к исследованию свойств новых версий сетей YOLO библиотеки Ultralytics возникает в том числе потому, что традиционно компания Ultralytics не выпускает сопровождающих статей при их релизе, но научное сообщество с энтузиазмом делает это за нее² [12–13].

Включение в список сравнительно старых НС PSPNet и U-Net позволит оценить, насколько далеко продвинулись новые сетевые модели и насколько результативен базовый подход семантической сегментации.

1 Jocher G., Qiu J., Chaurasia A. Ultralytics YOLO. [Электронный ресурс]. Режим доступа: <https://github.com/ultralytics/ultralytics> (дата обращения: 02.04.2025).

2 Ghosh A. YOLO11: Faster than you can imagine! [Электронный ресурс]. Режим доступа: <https://learnopencv.com/yolo11> (дата обращения: 15.10.2024).

Многочисленные исследования аугментации [14–16] свидетельствуют о том, что применение этой процедуры повышает качество обучения НС. Однако в данной работе настройка аугментации производилась таким образом, чтобы она обеспечивала равные условия обучения для всех исследуемых моделей НС. Поэтому для дополнения обучающего набора данных применялась только яркостная (изменение яркости, контраста, насыщенности, оттенка) и геометрическая (изменение размера, кадрирование, отражение по вертикали и горизонтали) коррекция изображений и не использовались настройки, доступные только в сетях YOLO, такие, например, как поворот изображения и объединение изображений (создание мозаики).

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

При обучении НС семантической сегментации были использованы базовые настройки планировщика schedule_160k.py. Однако число итераций было увеличено с 160 000 до 240 000. Это связано с тем, что при стандартном числе итераций некоторые НС оказались недообученными. Увеличение количества итераций обеспечило выход на плато всех анализируемых архитектур НС (на рис. 2 представлен пример графика изменения точности модели в процессе обучения). При этом в качестве оптимизатора для сетей DeepLabv3+, PSPNet, U-Net был использован алгоритм SGD, для сетей SegFormer, Twins и ConvNeXt — AdamW.

Для НС сегментации экземпляров количество эпох было установлено равным 240.

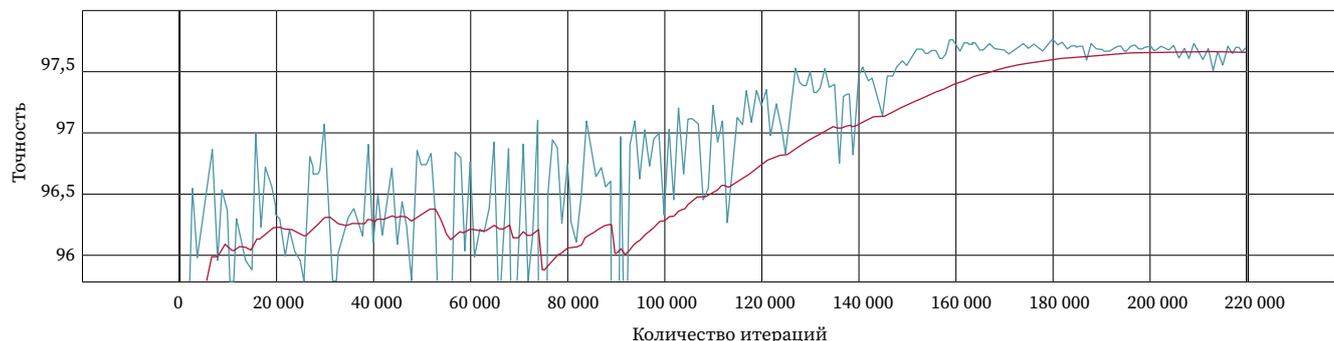
Валидация моделей производилась на тестовой выборке, которую составили 200 изображений.

Рис. 2 Ⓣ

График изменения точности модели в процессе обучения для НС PSPNet

Fig. 2

Training accuracy curve example for PSPNet



2.4 Метод сравнения качества обучения нейронных сетей различных типов

Стандартными показателями качества обучения для НС сегментации экземпляров являются метрики mAP50 и mAP50-95, для НС семантической сегментации — метрики accuracy и IoU. Однако сравнение результативности распознавания объектов сетями разных типов по соответствующим метрикам невозможно, т. к. эти метрики имеют принципиально разные алгоритмы расчета. Расчет mAP50 и mAP50-95 основывается на количестве правильно и ошибочно распознанных объектов, а также степени уверенности сети в отнесении объекта к данному классу (*англ.* confidence). НС семантической сегментации не разделяет результаты распознавания на отдельные объекты и не вычисляет параметр confidence, в результате 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. Пока этот процесс нельзя назвать завершенным: предлагаются различные подходы расчета показателей результативности распознавания как для семантической

сегментации³, так и для сегментации экземпляров [17]. Это обстоятельство позволяет предложить свой уникальный способ оценки качества обучения НС разных типов, который обеспечивал бы их сравнение.

Для сравнения НС разных типов между собой наиболее подходящей является метрика IoU, которая может быть рассчитана как при 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 различаются.

В задаче сегментации экземпляров метрика IoU представляет собой степень перекрытия эталонного и предсказанного контуров для каждого объекта в отдельности и может сильно изменяться у одной и той же обученной модели в зависимости от установленного порога confidence и значения подавления немаксимумов (*англ.* non-maximum suppression) — параметра, который отвечает за частоту построения перекрывающихся контуров распознаваемых объектов. При этом могут пересекаться границы объектов, принадлежащих не только к разным классам, но и к одному классу.

В НС семантической сегментации такие настройки, 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 возникают 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.
3. В случае перекрытия контуров объектов разных классов зона перекрытия рассматривается как особый сегмент. Принадлежность пикселей в этом особом сегменте к соответствующим классам считается равновероятной. Таким образом, при определении суммарных площадей истинно положительных (TP), ложноположительных (FP), ложноотрицательных (FN) сегментов площадь особого сегмента умножается на соответствующую вероятность, что учитывается в дальнейшем при расчете метрики IoU.

Последний случай в сформированном для эксперимента наборе данных не встречался, поэтому не представляется возможным определить, как учет такого эффекта влияет на конечные значения IoU.

Метрика IoU, таким образом, может быть рассчитана аналогично моделям семантической сегментации. При таком подходе теряется ключевая идея работы алгоритма сегментации экземпляров — разделение результатов предсказания на отдельные объекты внутри класса, но в то же время 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. Значения, полученные для НС YOLOv11, приведены в табл. 2. Закономерность изменения IoU для других НС семейства YOLO имеет схожий характер.

Анализ приведенных в табл. 2 данных позволяет определить оптимальные значения confidence и non-maximum suppression равными 0,1 и 0,5 соответственно.

³ Monteux A. Metrics for semantic segmentation. [Электронный ресурс]. Режим доступа: <https://ilmonteux.github.io/2019/05/10/segmentation-metrics.html> (дата обращения: 14.11.2024).

Таблица 2 ↗

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

Table 2

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

		Параметр non-maximum suppression									
		0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Параметр confidence	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

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

При выборе значения параметра non-maximum suppression очевидно, что приравнивание его к 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.

3 Результаты и обсуждение

3.1 Результаты обучения нейронных сетей

Стандартные показатели качества обучения НС, рассчитанные по тестовому набору данных, приведены в табл. 3 (для сетей сегментации экземпляров), табл. 4 (для сетей семантической сегментации). При этом метрики, представленные в табл. 3, рассчитывались для сегментов (не для ограничивающих рамок типа box) распознанных объектов.

Примеры результатов предсказаний на тестовых изображениях обученных моделей НС приведены на рис. 3.

Таблица 3 ↘

Показатели качества для обученных моделей сегментации экземпляров

Table 3

Metrics for trained instance segmentation models

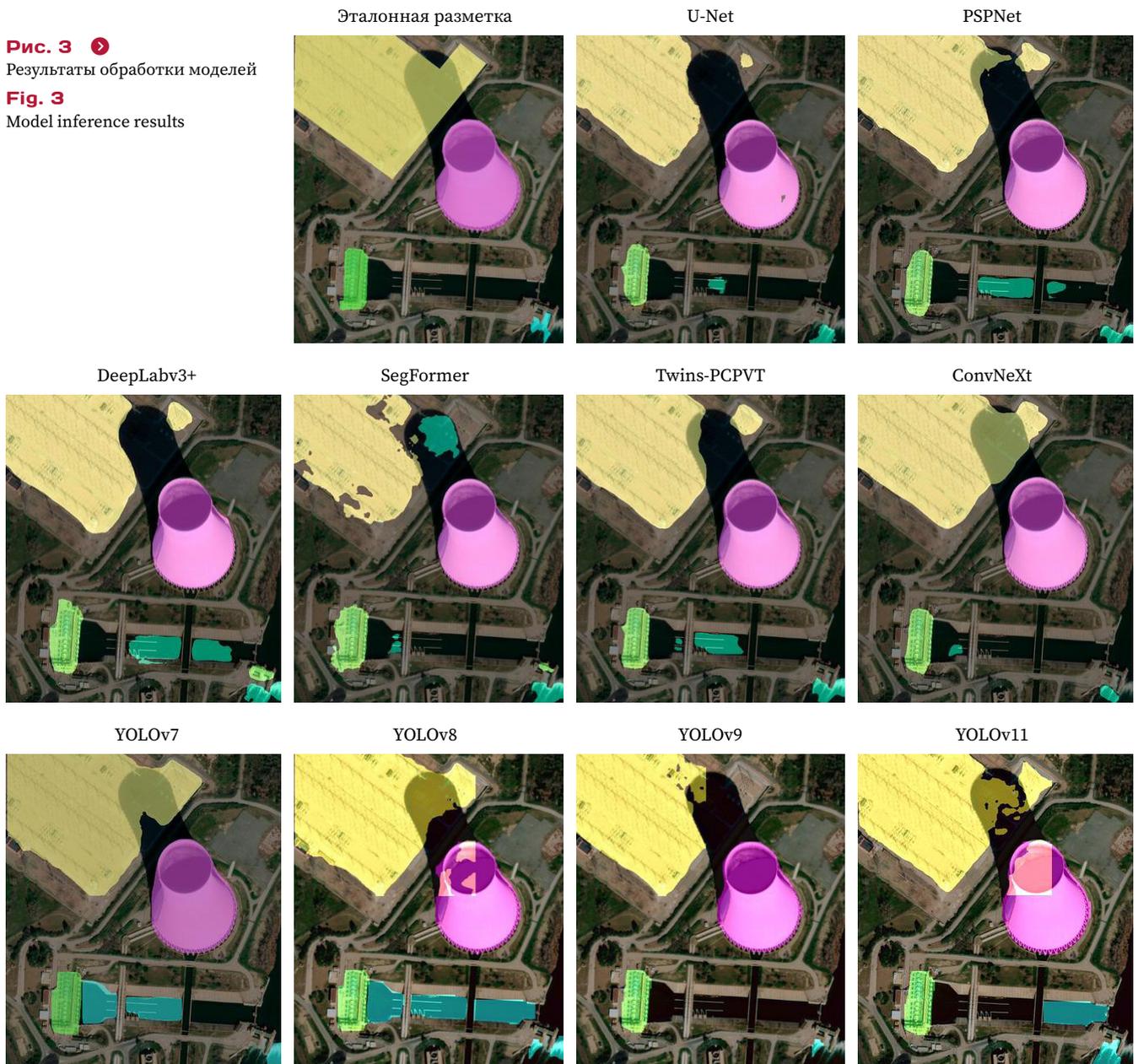
Модель	Градирни		Резервуары		РУ		БНС		ПЗ		РСВ		Среднее по всем классам	
	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95	mAP50	mAP50-95
YOLOv7	1,00	0,84	0,90	0,60	0,67	0,41	0,65	0,33	0,49	0,29	0,32	0,10	0,67	0,43
YOLOv8	0,99	0,84	0,91	0,59	0,77	0,47	0,83	0,47	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

Таблица 4 
Показатель точности для обученных моделей семантической сегментации

Table 4
Accuracy scores for trained semantic segmentation models

Модель	Градирни	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

Рис. 3 
Результаты обработки моделей
Fig. 3
Model inference results



Анализ значений показателей качества обучения НС и визуальный анализ результатов распознавания позволили сделать следующие выводы:

1. Отмечается корреляция между показателями качества обучения НС и степенью вариативности границ распознаваемых объектов: объекты РСВ имеют самые низкие показатели для обоих типов НС. Градирни, напротив, несмотря на меньшее по сравнению с РСВ количество меток, распознаются более уверенно.
2. На показатели качества обучения НС также влияют геометрические размеры распознаваемых объектов, что подтверждается работой [3].
3. На примере сегментации РУ (рис. 3) видно, что с задачей распознавания объекта, если он находится в тени, наиболее корректно 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.

3.2 Результаты сравнения нейронных сетей разных типов

Таблица 5 

Показатели качества обучения для апробированных моделей по метрике IoU

Table 5

Training quality metrics for evaluated models (IoU metric)

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 табл. 5.

Модель	Градирни	Резервуары	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

Анализ приведенных в табл. 5 данных позволил сделать следующие выводы.

1. При распознавании антропогенных объектов в целом 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 применительно к рассматриваемой задаче разница между максимальным значением 73,5, характерным для НС Twins-PCPVT, и значением 69,4 у НС ConvNeXt не принципиальна.
2. 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 третьем месте в наборе обучающих данных, ниже расположены только РСВ, имеющие на снимках устойчивые яркостные и текстурные характеристики, и градири — сравнительно большие объекты, имеющие устойчивые геометрические формы и размеры.

3. Крупные объекты с минимальной вариативностью (градири) 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.
4. Резервуары, являющиеся небольшими объектами с минимальной степенью вариативности границ, в целом с соразмеряемой достоверностью распознаются как НС семантической сегментации, так и НС сегментации экземпляров. Чуть более низкие показатели IoU у последних объясняются перекрытием областей расположения у близко размещенных по отношению друг к другу объектов.
5. Низкие показатели IoU при распознавании РСВ имеют несколько причин: самое высокое отношение среднеквадратичного отклонения / средней площади, низкая частота появления на снимках и выраженная вариативность контуров (т. е. самое малое значение совпадения контуров при повторной разметке).

4 Выводы

По результатам исследований можно сделать следующий вывод: 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, 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 необходимо исследовать применимость моделей НС к более широкому классу объектов с учетом степени вариативности их границ. Следует также оценить влияние гиперпараметров на улучшение обучения моделей. Кроме того, большой интерес представляет анализ методов паноптической сегментации, которые на данном этапе авторами не рассматривались.

БЛАГОДАРНОСТИ

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

БИБЛИОГРАФИЯ

1. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. P. 3431–3440. DOI:10.1109/CVPR.2015.7298965.
2. He K., Gkioxari G., Dollar P., et al. Mask R-CNN // Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017. P. 2980–2988. DOI:10.1109/ICCV.2017.322.
3. Zhao H., Puig X., Xiao T., et al. Semantic Understanding of Scenes through the ADE20K Dataset // Preprint arXiv.org, 2016. [Электронный ресурс]. Режим доступа: <https://arxiv.org/pdf/1608.05442> (дата обращения: 11.07.2024).

4. Ronneberger O., Fischer Ph., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Preprint arXiv.org, 2015. [Электронный ресурс]. Режим доступа: <https://arxiv.org/pdf/1505.04597> (дата обращения: 11.07.2024).
5. Zhao H., Shi J., Qi X., et al. Pyramid scene parsing network // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. P. 6230–6239. DOI:10.1109/CVPR.2017.660.
6. Chen L.C., Zhu Y., Papandreou G., et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation // Proceedings of 15th European Conference “Computer Vision – ECCV 2018”, Munich, September 8–14, 2018. Cham: Springer, 2018. P. 833–851. DOI:10.1007/978-3-030-01234-2_49.
7. Xie E., Wenhai W., Zhiding Y. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers // Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: Red Hook, 2021. P. 12077–12090.
8. Chu X., Tian Z., Wang Y., et al. Twins: Revisiting the Design of Spatial Attention in Vision Transformers // Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: Red Hook, 2021. P. 9355–9367.
9. Liu Z., Mao H., Wu C.-Y., et al. A ConvNet for the 2020s // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. P. 11966–11976. DOI:10.1109/CVPR52688.2022.01167.
10. Wang C.Y., Bochkovskiy A., Liao H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. P. 7464–7475. DOI:10.1109/CVPR52729.2023.00721.
11. Wang Ch.-Y., Yeh I.H., Mark Liao H.-Y. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information // Proceedings of 18th European Conference “Computer Vision – ECCV 2024”, Milan, September 29 – October 4, 2024. Cham: Springer, 2024. P. 1–21. DOI:10.1007/978-3-031-72751-1_1.
12. Sohan M., Sai Ram T., Rami Reddy Ch.V. A Review on YOLOv8 and its Advancements // Proceedings of the International Conference on Data Intelligence and Cognitive Informatics. Singapore: Springer, 2024. P. 529–545. DOI:10.1007/978-981-99-7962-2_39.
13. Terven J., Cordova-Esparza D.-M., Romero-González J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS // MAKE. 2023. Vol. 5. No. 4. P. 1680–1716. DOI:10.3390/make5040083.
14. Shorten C., Khoshgoftaar T.M. A survey on Image Data Augmentation for Deep Learning // Journal of Big Data. 2019. Vol. 6. P. 60. DOI:10.1186/s40537-019-0197-0.
15. Zoph B., Cubuk E.D., Ghiasi G., et al. Learning data augmentation strategies for object detection // Preprint arXiv.org, 2019. [Электронный ресурс]. Режим доступа: <https://arxiv.org/pdf/1906.11172> (дата обращения: 29.04.2024).
16. Ghaffar M.A.A., McKinstry A., Maul T., 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. DOI:10.5194/isprs-annals-IV-2-W7-47-2019.
17. Chen L., Wu Y., Stegmaier J., et al. SortedAP: Rethinking evaluation metrics for instance segmentation // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Paris: IEEE, 2023. P. 3925–3931. DOI:10.1109/ICCVW60793.2023.00424.

АВТОРЫ Саидов Алишер Голибович

ФГБВОУ ВО «Военно-космическая академия имени А.Ф. Можайского»,
 Санкт-Петербург, Россия
 33-й отдел, Военный институт (научно-исследовательский)

Жуков Денис Валерьевич

ФГБВОУ ВО «Военно-космическая академия имени А.Ф. Можайского»,
 Санкт-Петербург, Россия
 33-й отдел, Военный институт (научно-исследовательский)
 канд. техн. наук

Поступила 09.04.2025. Принята к публикации 22.08.2025. Опубликовано 29.08.2025.



Features of neural network application in recognizing anthropogenic objects with variable contours

Alisher G. Saidov¹, Denis V. Zhukov¹

¹ Mozhaisky Military Space Academy, St. Petersburg, Russia
 spb_pilligrim83@mail.ru

CITATION Saidov A.G., Zhukov D.V. Features of neural network application in recognizing anthropogenic objects with variable contours. *Izvestia vuzov. Geodesy and Aerophotosurveying*. 2025;69(4): 83–95. DOI:10.30533/GiA-2025-030.

KEYWORDS semantic segmentation, instance segmentation, aerospace imagery, metrics, deep learning, dataset

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 authors categorize object contour variability into three levels and examine its impact on model accuracy and generalizability. A key focus of the study is the relationship between contour variability and the effectiveness of deep learning approaches. 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 paper also discusses segmentation quality metrics and their interpretation nuances. 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. In conclusion, the authors emphasize the critical role of contour variability in data preparation and segmentation method selection. They highlight the need for further research to enhance model training and improve object detection reliability.

ACKNOWLEDGEMENTS The authors express their special gratitude to Roman Khabarov for the inspiration and ideas that made this research possible.

REFERENCES 1. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE; 2015: 3431–3440. DOI:10.1109/CVPR.2015.7298965.

2. He K, Gkioxari G, Dollar P, et al. Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE; 2017: 2980–2988. DOI:10.1109/ICCV.2017.322.
3. Zhao H, Puig X, Xiao T, et al. Semantic Understanding of Scenes through the ADE20K Dataset. *Preprint arXiv.org*, 2016. Available from: <https://arxiv.org/pdf/1608.05442> (Accessed 11 July 2024).
4. Ronneberger O, Fischer Ph, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Preprint arXiv.org*, 2015. Available from: <https://arxiv.org/pdf/1505.04597> (Accessed 11 July 2024).
5. Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE; 2017: 6230–6239. DOI:10.1109/CVPR.2017.660.
6. Chen LC, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of 15th European Conference “Computer Vision – ECCV 2018”, Munich, September 8–14, 2018*. Cham: Springer; 2018: 833–851. DOI:10.1007/978-3-030-01234-2_49.
7. Xie E, Wenhai W, Zhiding Y. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. New York: Red Hook; 2021: 12077–12090.
8. Chu X, Tian Z, Wang Y, et al. Twins: Revisiting the Design of Spatial Attention in Vision Transformers. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. New York: Red Hook; 2021: 9355–9367.
9. Liu Z, Mao H, Wu CY, et al. A ConvNet for the 2020s. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE; 2022: 11966–11976. DOI:10.1109/CVPR52688.2022.01167.
10. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE; 2023: 7464–7475. DOI:10.1109/CVPR52729.2023.00721.
11. Wang ChY, Yeh IH, Mark Liao HY. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *Proceedings of 18th European Conference “Computer Vision – ECCV 2024”, Milan, September 29 – October 4, 2024*. Cham: Springer; 2024: 1–21. DOI:10.1007/978-3-031-72751-1_1.
12. Sohan M, Sai Ram T, Rami Reddy ChV. A Review on YOLOv8 and its Advancements. *Proceedings of the International Conference on Data Intelligence and Cognitive Informatics*. Singapore: Springer; 2024: 529–545. DOI:10.1007/978-981-99-7962-2_39.
13. Terven J, Cordova-Esparza DM, Romero-González JA. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*. 2023;5(4): 1680–1716. DOI:10.3390/make5040083.
14. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*. 2019;6: 60. DOI:10.1186/s40537-019-0197-0.
15. Zoph B, Cubuk ED, Ghiasi G, et al. Learning data augmentation strategies for object detection. *Preprint arXiv.org*, 2019. Available from: <https://arxiv.org/pdf/1906.11172> (Accessed 29 April 2024).
16. Ghaffar MAA, McKinstry A, Maul T, et al. Data augmentation approaches for satellite image super-resolution. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2019;IV-2/W7: 47–54. DOI:10.5194/isprs-annals-IV-2-W7-47-2019.
17. Chen L, Wu Y, Stegmaier J, et al. SortedAP: Rethinking evaluation metrics for instance segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. Paris: IEEE; 2023: 3925–3931. DOI:10.1109/ICCVW60793.2023.00424.

AUTHORS **Alisher G. Saidov**

Mozhaisky Military Space Academy, St. Petersburg, Russia
33rd Department, Military Institute (Research)

Denis V. Zhukov

Mozhaisky Military Space Academy, St. Petersburg, Russia
33rd Department, Military Institute (Research)
PhD in Engineering

Submitted: April 09, 2025. Accepted: August 22, 2025. Published: August 29, 2025.