# Модификация нейронной сети U-Net в задаче анализа изображений с БПЛА

Р. В. Ларионов<sup>1</sup>, Н. В. Котов<sup>2</sup>

Ярославский государственный университет им. П. Г. Демидова <sup>1</sup>r.larionov3@uniyar.ac.ru, <sup>2</sup>n.kotov1@uniyar.ac.ru

Аннотация. Представлены результаты применения методов машинного обучения в задаче бинарной сегментации изображений с беспилотных летательных аппаратов для обнаружения незаконных свалок твердых бытовых отходов. В качестве алгоритмов сегментации рассмотрены сверточные нейронные сети на основе архитектуры U-Net и Deeplabv3. База изображений для обучения содержит снимки 87 свалок на территории Ярославской области. Обучение И тестирование алгоритмов проведено на суперкомпьютере NVIDIA DGX-1. Качество работы алгоритмов оценивалось с использованием коэффициента Серенсена и метрики F1. Наилучшие значения коэффициента Серенсена и метрики F1 составили 0,45 и 0,23 соответственно.

Ключевые слова: сверточная нейронная сеть, сегментация изображения, компьютерное зрение, свалки

## I. Введение

Незаконная утилизация отходов является одним из наиболее важных проблем на сегодняшний день, которая способствует нарастанию социальных и экологических волнений, особенно в развивающихся странах [1]. Незаконная утилизация отходов угрожает общественной безопасности и здоровью, окружающей среде и государственной экономике. Часто происходят поджоги отходы, в результате чего выделятся высокотоксичные пары (например, диоксин), а также возможно возникновение неконтролируемых лесных пожаров, что было особенно актуально в 2021 году [2].

На несанкционированных свалках часто отсутствует надлежащая обработка отходов, что приводит к выбросу фильтрата, который загрязняет источники воды и наносит долговременный ущерб, например, увеличивая заболеваемость раком населения [3]. В [4] авторы влияние незаконных перечисляют свалок на окружающую среду, которые включают загрязнение растений, животных и воздуха, почвы, воды, что вызывает нейротоксичность, инфекционные заболевания и респираторные заболеваний у людей. Обнаружение незаконных свалок имеет решающее значение для предотвращения и уменьшения их воздействия на окружающую среду и уменьшения стоимости борьбы с их устранением.

Хотя в последние годы были предпринято множество усилий по обнаружению подозрительных областей захламления на изображениях, ручная обработка больших потоков информации до сих пор является доминирующим способом обнаружения незаконных мест складирования отходов. Регулярный анализ территории затруднен необходимостью привлечения огромных человеческих ресурсов. Последние достижения в методах компьютерного зрения, в основном полученные с использованием моделей глубокого обучения делают возможным обнаружение свалок в автоматическом режиме с точностью, не уступающей человеческой [5].

Соответственно, целью данного исследования является разработка алгоритма семантической сегментации свалок твердых бытовых отходов по данным БПЛА с применением алгоритмов глубокого обучения.

#### II. РАЗРАБОТКА АЛГОРИТМА СЕГМЕНТАЦИИ

В рамках данной работы проведена разработка алгоритма сегментации свалок твердых бытовых отходов на изображениях БПЛА. Для этого поставлено несколько задач:

- сбор и формирование базы данных для обучения и тестирования алгоритма, аугментация данных;
- разработка архитектуры сверточной нейронной сети для сегментации объектов на изображении;
- обучение и тестирование сверточной нейронной сети;
- исследование по оценке точности автоматической сегментации объектов в зависимости от различных параметров – архитектуры нейронной сети, функции потерь, оптимизатора.

## А. База изображений

Сбор базы данных изображений для обучения и тестирования алгоритма сегментации произведен с использованием БПЛА DJI Mavic Air. Съемка осуществлялась в различное время суток с высотой съемки от 30 метров до 100-230 метров в зависимости от внешних факторов с шагом 10 метров. Всего получено 1114 изображений, размером 4056 × 3040 пикселей, из которых случайным образом отобраны по 2 изображения для каждого объекта. Всего свалок 87. Таким образом, база данных состоит из 174 изображений.

Далее исходные изображения разделены на прямоугольные области (патчи) методом скользящего окна размером 512×512 пикселей с шагом 256 пикселей [6].

ТАБЛИЦА I СТАТИСТИКА СФОРМИРОВАННОЙ БАЗЫ ИЗОБРАЖЕНИЙ

Тип патчей	Подвыборки		
	Обучающая	Тестовая	
Всего	2250	612	
С объектами	495	153	
Без объектов	1755	459	

Сформированная база разделена на тренировочную и тестовую подвыборки как в отношении 80:20. При этом 10% от обучающей подвыборки выделено на валидационный набор, который необходим для контроля процесса обучения нейронной сети. Все изображения приведены в диапазон [0;1] путем деления на 255 – максимальное значения яркости сенсора БПЛА. Подробная статистика по сформированной базе изображений представлена в табл. 1.

Для искусственного увеличения количества изображений проведена аугментация данных, состоящая из трех этапов [7]:

- повороты изображений на 90, 180, 270 градусов, зеркальные отражения изображений;
- хроматические искажения. Изображения переводились из цветовой модели RGB (красный, зеленый, синий) в HSV (тон, насыщенность, значение). Затем в базисе HSV к значениям тона прибавлялись случайные числа в пределах [-100; 100], а к значениям насыщенности и яркости прибавлялись случайные числа из отрезка [-0.02; 0,02]. Если значение яркости выходило за пределы [0; 1], то такие значения усекались до 0 или до 1 соответственно;
- случайные сдвиги, растяжения до 20% от исходного размера изображения и повороты фрагментов на небольшой угол в пределах от -15 до +15 градусов.

# В. Обучение нейронной сети

В рамках данной работы исследовались три сверточные нейронные сети – U-Net [8] + ResNet-34 [9], U-Net + SE-ResNeXt-50 [10], DeeplabV3 [11] + Xception, которые названы U-ResNet-34, U-SE-ResNeXt-50 и DeeplabV3+. Для обучения U-ResNet-34 использовалась одна видеокарта NVIDIA Tesla V100 с 16 Гб видеопамяти, а для обучения U-SE-ResNeXt-50, DeeplabV3plus были задействованы две видеокарты.

Обучение нейронной сети производилось на протяжении 100 эпох с размером батча 16. В качестве функции потерь использована бинарная кросс-энтропия плюс коэффициент Серенсена, которая вычисляется по формуле:

$$Loss = -\sum_{x,y} \left( x \log(y) + (1-x) \log(1-y) \right) + dice$$
(1)

ТАБЛИЦА II РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ

Mornuuu	Алгоритмы		
метрики	U-ResNet-34	U-ResNet-50	DeeplabV3plus
F1-score	0,18	0,23	0,17
Dice	0,31	0,45	0,30

ТАБЛИЦА III ТЕСТИРОВАНИЕ U-RESNET-50 С РАЗЛИЧНЫМИ ФУНКЦИЯМИ ПОТЕРЬ

Метрики	Функция потерь		
	BCE_dice_loss	Focal_loss	
F1-score	0,23	0,21	
Dice	0,45	0,42	

ТАБЛИЦА IV ТЕСТИРОВАНИЕ U-RESNET-50 С РАЗЛИЧНЫМИ ОПТИМИЗАТОРАМИ

Метрики	Оптимизаторами			
	Adam	RMSProp	SGD	
F1-score	0,23	0,23	0,14	
Dice	0,45	0,44	0,29	

Для оценки работы разработанного алгоритма использовались коэффициент Серенсена (dice) и метрика F1-score.

Результаты тестирования всех трех алгоритмов представлены в табл. 2.

Согласно результатам, из табл. 2 архитектура нейронной сети U-ResNet-34 показала результат немного лучше, чем архитектура DeeplabV3+, но проигрывает по всем метрикам нейронной сети U-SE-ResNeXt-50: значение метрики F1-score у новой модели составляет 0,23, тогда как у U-ResNet-34 данный показатель не превышает значения 0,18, а значение метрики Dice возросло до 0,45.

## С. Оценка точности в зависимости от функции потерь

Рассмотрим влияние функций потерь на качество точности работы для нейронной сети U-SE-ResNeXt-50. В качестве функции потерь рассмотрены две спецификации: бинарная кросс-энтропия плюс один минус коэффициент Серенсена (BDE\_dice\_loss) И функция потерь (Focal\_loss) фокальная [10] с параметрами  $\alpha = 0.25$  и  $\gamma = 2$ . Результаты тестирования представлены в табл. 3.

Как видно из табл. 3, обучение с использованием функции потерь BCE\_dice\_loss показало лучший результат: показатели метрики Dice выше на 0,03, а метрика F1-score больше на 0,02. Разницу можно объяснить тем, что для обучения с использованием фокальной функции потерь необходимо использовать другой алгоритм оптимизации, а также уменьшить коэффициент обучения: значения функции в процессе обучения уменьшается до 0,2 до 0,0002, тогда как значение BCE\_dice\_loss уменьшается с 1,5 до 0,12. Из разницы порядков в значениях функции следует, что для обучения с использованием Focal\_loss значение коэффициента обучения, которое влияет на скорость изменения градиента в ходе градиентного спуска, 0,001 велико.



a)

Рис. 1. Результаты работы нейронной сети U-ResNet-50: а) оригинальное изображение; б) результат сегментации (белым цветом выделены границы свалки)

# D. Оценка точности в зависимости от различных оптимизаторов

Рассмотрим влияние алгоритма оптимизации процесса обучения сверточных нейронных сетей на процесс обучения алгоритма U-SE-ResNeXt-50. В качестве алгоритмов оптимизации рассмотрены: Adam, SGD RMSProp, [11]. Результаты тестирования представлены в табл. 4.

Согласно результатам табл. 4, алгоритм оптимизации SGD значительно проигрывает алгоритмам Adam и которые показывают близкие RMSProp, очень результаты, однако Adam демонстрирует незначительное превосходство.

На рис. 1 представлен пример результата работы нейронной сети U-SE-ResNeXt-50 на тестовых изображениях.

## III. ЗАКЛЮЧЕНИЕ

В работе произведены исследования по разработке алгоритмов автоматического сегментирования свалок твердых бытовых отходов на изображениях с БПЛА. По результатам тестирования наилучший показатель со значениями метрик Dice и F1-score, равными 0,45 и 0,23 соответственно, показала архитектура нейронной сети U-SE-ResNeXt-50. Использование функции потерь BCE\_dice\_loss предпочтительнее, чем использование Focal\_loss. Выполнено сравнение трех алгоритмов оптимизации процесса обучения нейронной сети: Adam, RMSProp, SGD. Использование Adam показало лучшее значение метрики Dice на 0,01 по сравнению с RMSProp и на 0.15 по сравнению с SGD. Полученные результаты позволяют утверждать, что сверточные нейронные сети могут использоваться для обнаружения незаконных мусорных полигонов на спутниковых изображениях, но алгоритмы требуют дальнейших исследований по улучшению качества работы.

#### Список литературы

- [1] Shrivastava P., Mishra S., Katiyar S. K. A Raview of Solid Waste Management Techniques Using GIS and Other Technologies // 2015 International Conference on Computational Intelligence and Communication Network (CICN). 2015. P. 1456-1459.
- Rocco G. Survival After Surgical Treatment of Lung Cancer Arising [2] in the Population Exposed to Illegal Dumping of Toxic Waste in the Land of Fires of Southern Italy // Anticancer Res., vol. 36, no 5. 2016. Pp. 2119-2224.
- [3] Schrab G.E., Brown K.W., Donnelly K.C. Acute and genetic toxicity of municipal landfill leachate // Water Air Soil Pollut., vol. 69. 1993. Pp. 99-112.
- [4] Limoli A., Garzia E., De Pretto A., De Muri C. Illegal landfill in Italy (EU) - a multidisciplinary approach // Environmental Forensics, vol. 20, no 1. 2019. Pp. 26-38.
- Youme O., Bayet T., Dembele J., Cambier C. Deep Learning and [5] Remote Sensing: Detection of Dumping Waste Using UAV // Procedia Computer Science, vol. 185. 2021. Pp. 361-369.
- Ларионов Р.В., Павлов В.А., Хрящев В.В. Использование [6] изображений с БПЛА для обнаружения незаконных свалок твердых бытовых отходов // Цифровая обработка сигналов и ее применение (DSPA-2021): докл. 23-й междунар. Конф., Москва, 2021. C. 209-213.
- [7] Khryashchev V., Larionov R. Wildfire Segmentation on Satellite Images using Deep Learning // 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), Moscow, Russia. 2020. Pp. 1-5. Doi: 10.1109/MWENT47943.2020.9067475
- Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks [8] for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Lecture Notes in Computer Science. Springer. Vol. 9351. 2015. P. 234-241.
- He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image [9] Recognition // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770-778.
- [10] Jie H., Li Sh., Gang S. Squeeze-and-Excitation Networks // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. Doi: 10.1109/CVPR.2018.00745.
- [11] Chen L.-Ch., Zhu Y., Papandreou G., Florian S., Hartwing A. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation // Computer Vision and Pattern Recognition. arXiv:1802.02611, 2018.