

Способ семантической сегментации тепловизионных изображений

А.В. Мингалев¹, А.В. Белов¹, С.Н. Шушарин¹, М.П. Шлеймович²

¹ *Научно-производственное объединение «Государственный институт прикладной оптики», Казань*

² *Казанский национальный исследовательский технический университет им. А. Н. Туполева – КАИ*

Аннотация: Представлены результаты исследования, направленного на разработку способа семантической сегментации тепловизионных изображений с применением модифицированного нейросетевого алгоритма, отличающегося от исходного нейросетевого алгоритма более высокой скоростью обработки графической информации. В рамках исследования выполнена модификация нейросетевого алгоритма семантической сегментации DeepLabv3+ за счет уменьшения количества параметров нейросетевой модели, что позволило увеличить скорость обработки графической информации на 48% с 27 до 40 кадров в секунду. Также представлена методика обучения, позволяющая повысить точность модифицированного нейросетевого алгоритма, при этом получено значение точности ниже точности исходного нейросетевого алгоритма на 5%. Скорость обработки графической информации модифицированным нейросетевым алгоритмом, составляющая 40 кадров в секунду, позволяет применять данный алгоритм для дешифрирования графической информации в составе высокопроизводительных съемочных систем в режиме реального времени непосредственно в процессе выполнения съемочных работ при решении задач в лесном хозяйстве, задач экологического мониторинга, мониторинга стихийных бедствий.

Ключевые слова: нейросетевые алгоритмы, семантическая сегментация, машинное обучение, аугментация данных.

Описание исследования

В настоящий момент существует множество методов обработки графической информации при выполнении дешифрирования, реализующих различные подходы для обнаружения объектов на изображениях. На текущий момент развития методы на основе сверточных нейронных сетей (далее СНС) являются наиболее перспективным направлением для решения научной задачи обнаружения объектов на изображениях.

В процессе дешифрирования тепловизионных изображений, формируемых авиационными съемочными системами при решении задач в лесном хозяйстве, задач экологического мониторинга, мониторинга стихийных бедствий, необходимо выполнять обнаружение и локализацию объектов (определять пиксельные координаты) с целью дальнейшего вычисления их

географических координат на местности. В связи с этим из известных типов алгоритмов на основе СНС [1], представленных в таблице № 1, наиболее подходящими для решения указанной задачи являются детектирующие нейросетевые алгоритмы и нейросетевые алгоритмы семантической сегментации. Эти алгоритмы не только позволяют дать ответ относительно наличия либо отсутствия на анализируемых кадрах объектов заданных типов, но и предоставляют информацию об их размерах и пиксельных координатах.

Таблица № 1

Характеристики различных типов нейросетевых алгоритмов

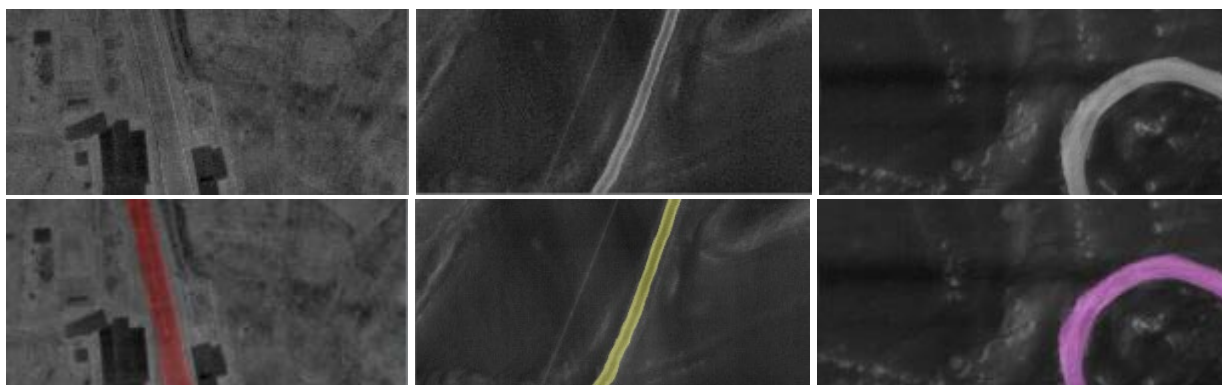
Типы алгоритмов	Назначение	Формат представления результата обработки	Скорость обработки	Наиболее известные алгоритмы
Классифицирующие	Обнаружение на изображении заданных типов классов объектов	Список со значениями вероятности наличия на изображении объектов заданных классов без локализации положения объекта на изображении	Высокая	lenet, alexnet, vgg, googlenet
Детектирующие	Обнаружение на изображении и локализация размеров и положения объектов	Список координат, вероятность и размеры прямоугольников, ограничивающих обнаруженные объекты заданных классов	Высокая	SSD, YOLO
Семантическая сегментация	Определение принадлежности каждого пикселя к одному из заданных классов	Изображение, равное по размеру исходному, пиксели, отнесенные алгоритмом к одному из классов объектов закрашиваются определенным цветом	Низкая	Unet, Enet, DeepLab

В рамках данного исследования рассмотрен нейросетевой алгоритм семантической сегментации DeepLabv3+ [2, 3] (далее DeepLab).

На первом этапе исследования выполнена оценка нейросетевого алгоритма DeepLab с целью определения скорости обработки графической информации и точности выполнения семантической сегментации данным алгоритмом. Для этого создан набор данных из тепловизионных изображений общим количеством 1100 изображений размером 576 на 576 пикселей.

Оператором выполнена разметка всех тепловизионных изображений из набора данных, при этом на изображениях были выделены объекты пяти классов:

- «Roads» - дороги с ровными контурами с темным оттенком (Рис. 1 а);
- «Track» - проселочные дороги с двумя полосами от колес (Рис. 1 б);
- «Country Road» - проселочные дороги, не имеющие асфальтного покрытия или проходящие через населенные пункты (Рис. 1 в);
- «River» - реки со светлым оттенком в виде тонкой полосы (Рис. 2 а);
- «Lake» - озера – объекты, имеющие характерный светлый ровный оттенок, которые занимают большую по сравнению с объектами «River» площадь изображения (Рис. 2 б).

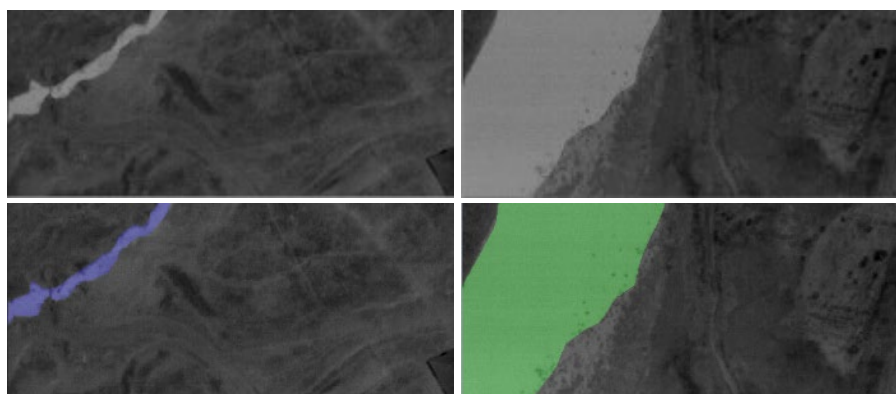


а) «Road»

б) «Track»

в) «CountryRoad»

Рис. 1. - Пример исходных изображений - верхний ряд, изображения из верхнего ряда с выполненной разметкой - нижний ряд



а) «River»

б) «Lake»

Рис. 2. - Пример исходных изображений – верхний ряд, изображения из верхнего ряда с выполненной разметкой - нижний ряд

Для обучения использована предобученная модель [4], в качестве библиотеки машинного обучения применена библиотека TensorFlow [5], а также исходный код программной библиотеки DeepLab [6], представленный в репозитории библиотеки TensorFlow. Выполнены тесты для оценки влияния различных параметров, таких как «Atrous rate», «Output stride» [2, 6] нейросетевого алгоритма DeepLab, задаваемых при обучении, на точность семантической сегментации и скорость обработки графической информации. Для всех вариантов параметров обучение выполнялось на протяжении 120000 итераций. Параметр режима обучения «FINE_TUNE_BATCH_NORM» установлен в значение «false». Полученные результаты представлены в таблице № 2. При этом точность семантической сегментации (далее – точность) вычислялась по формуле (1) [7].

Таблица № 2
Результаты оценки точности при различных параметрах

Параметр		Точность для различных классов, P, отн. ед.					Средняя точность, P, отн. ед.	Скорость, кадров/с
		«Roads»	«Track»	«Country Road»	«River»	«Lake»		
Atrous_rate	Output stride							
{3;6;9}	32	0,65	0,65	0,65	0,39	0,94	0,65	31
{6;12;18}	16	0,88	0,80	0,84	0,61	0,96	0,82	27
{12;24;36}	8	0,70	0,74	0,74	0,39	0,94	0,70	9

$$P = \frac{T_p}{T_p + F_p + F_n}, \quad (1)$$

где P – точность, T_p – количество верно определенных пикселей изображения, F_p – количество неверно определенных пикселей, F_n – количество не найденных пикселей, заданных в разметке.

В выполненных тестах наибольшее значение точности (таблица № 2) получено для нейросетевого алгоритма DeepLab с параметрами: «Atrous rate»: AR = {6; 12; 18} и «Output stride»: OS = 16, при этом скорость обработки составила 27 кадров в секунду, соответственно время обработки одного кадра составило 0,036 с.

На втором этапе исследования проведен анализ структуры нейросетевого алгоритма DeepLab и выполнена модификация данного алгоритма с целью исследования возможности повышения скорости обработки графической информации без снижения точности.

Структурно нейросетевой алгоритм семантической сегментации DeepLab состоит из кодера (encoder), который отвечает за извлечение признаков и уменьшение размеров карт признаков, и декодера (decoder) – блока, который увеличивает разрешение карт признаков до размеров исходного изображения (Рис. 3 [2]).

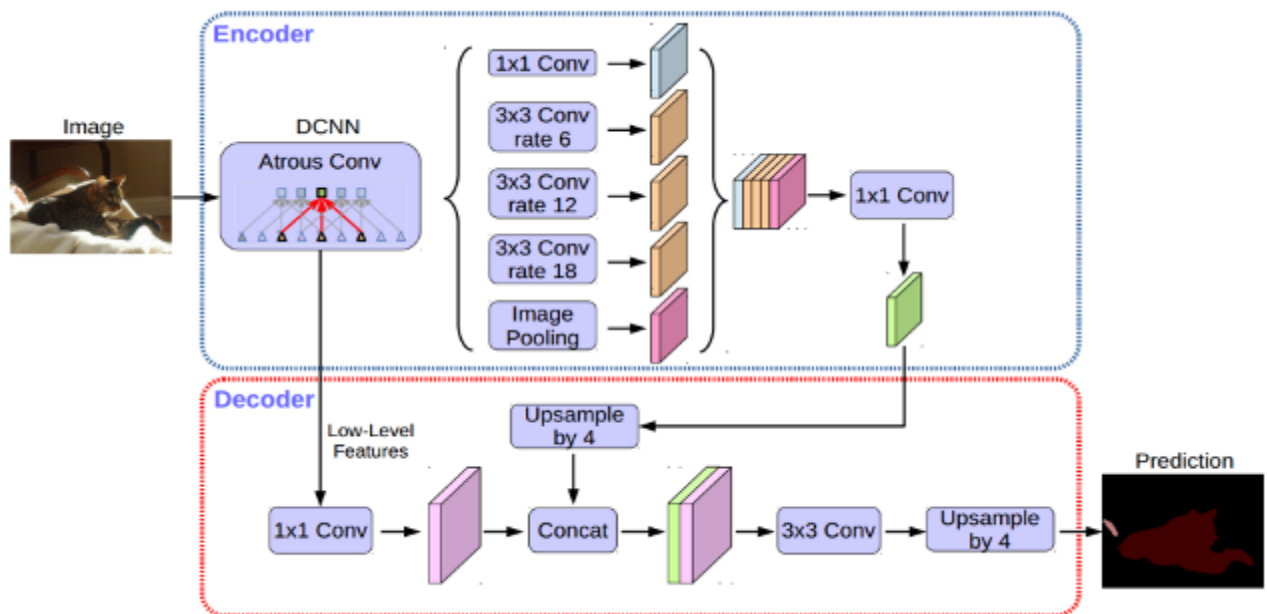


Рис. 3. - Структура нейросетевого алгоритма DeepLab

Кодер нейросетевого алгоритма DeepLab построен на основе сверточной нейронной сети Xception (Рис. 4). В свою очередь СНС Xception состоит из трех основных модулей: входной модуль - «Entry flow», средний модуль - «Middle flow» и выходной модуль - «Exit flow». Модуль «Middle flow» в свою очередь состоит из 16 блоков «Sep Conv 728, 3x3», а каждый блок «Sep Conv 728, 3x3» состоит из трех повторяющихся одинаковых групп слоев, включающих слой активации, слой раздельной свертки по глубине,

слой пакетной нормализации, сверточный слой, слой пакетной нормализации (Рис. 5).

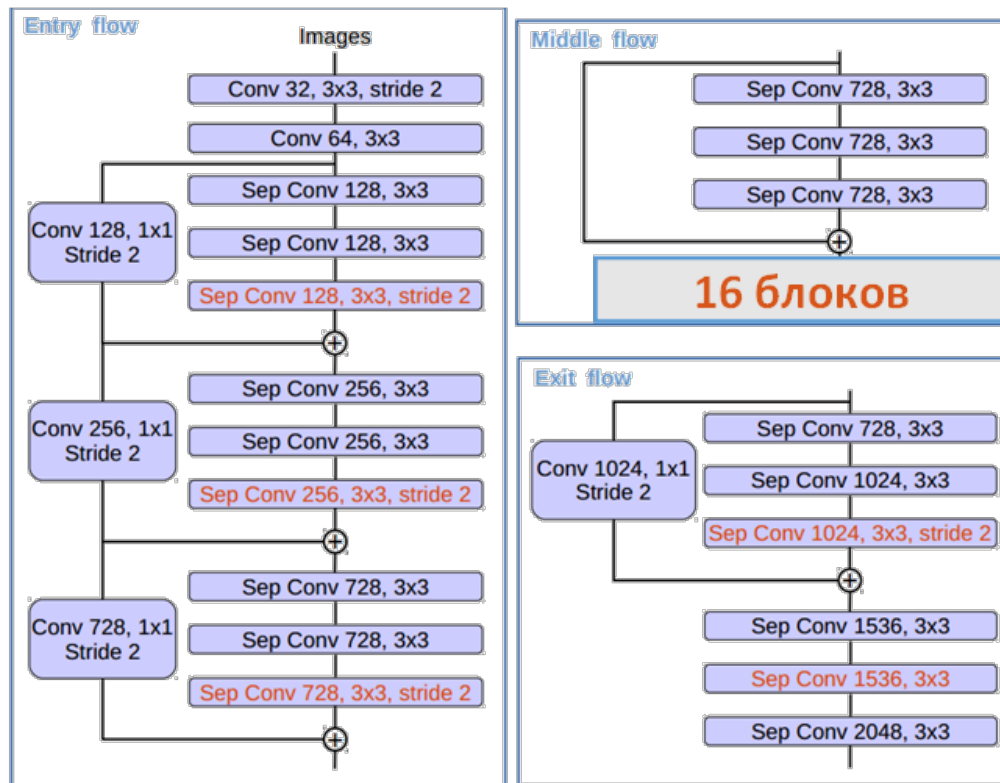


Рис. 4. - Структура СНС Xception

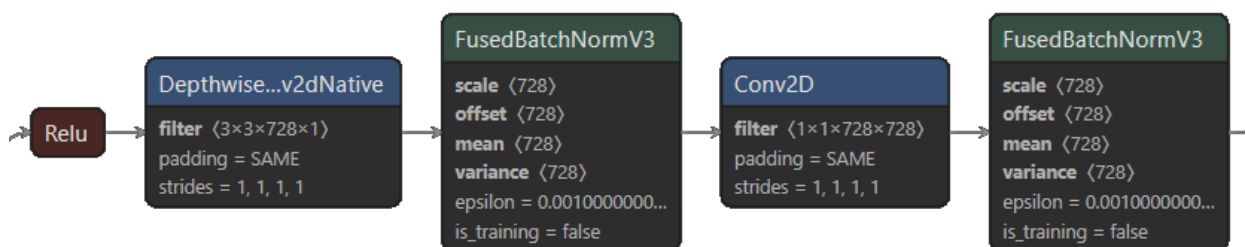


Рис. 5. - Группа слоев блока «Sep Conv 728, 3x3»

В рамках исследования описанная структура СНС Xception в кодере DeepLab с целью исследования возможности повышения скорости обработки графической информации была преобразована таким образом, чтобы Модуль «Middle flow» состоял не из 16, а из 4 блоков «Sep Conv 728, 3x3», в результате из нейросетевого алгоритма DeepLab получен модифицированный нейросетевой алгоритм за счет уменьшения количества весов нейросетевой

модели кодера на основе СНС Xception (Рис. 6). Модифицированный нейросетевой алгоритм обозначим как DL_Xception_4.

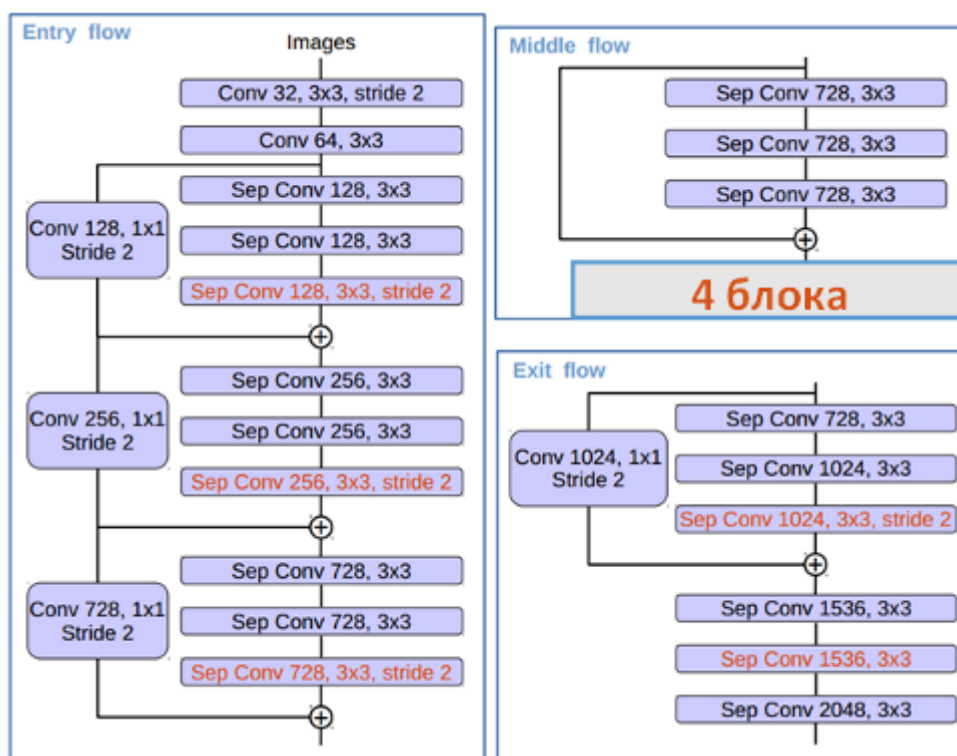


Рис. 6. - Структура модифицированной СНС Xception в кодере нейросетевого алгоритма DL_Xception_4

На следующем этапе исследования выполнена оценка модифицированного нейросетевого алгоритма DL_Xception_4 с целью определения скорости обработки графической информации и точности выполнения семантической сегментации данным алгоритмом.

Для обучения модифицированного нейросетевого алгоритма DL_Xception_4 использована предобученная модель [4], а также набор данных из тепловизионных изображений общим количеством 1100 изображений размером 576 на 576 пикселей, описанный ранее. Обучение выполнялось на протяжении 120000 итераций. Параметр режима обучения «FINE_TUNE_BATCH_NORM» установлен в значение «false». Полученные результаты представлены в таблице № 3.

Таблица № 3

Результаты оценки точности и скорости нейросетевых алгоритмов DeepLab и DL Xception 4

Метод		DeepLab	DL_Xception_4
Параметр			
Количество обучаемых параметров		40 949 638	21 581 926
Точность семантической сегментации, P	Класс «Road»	0,88	0,76
	Класс «Track»	0,80	0,77
	Класс «Country Road»	0,84	0,75
	Класс «Lake»	0,96	0,96
	Класс «River»	0,61	0,48
Средняя точность сегментации, P		0,82	0,74
Время обработки одного кадра, с		0,036	0,025

По результатам выполненных тестов у модифицированного нейросетевого алгоритма DL_Xception_4 по отношению к нейросетевому алгоритму DeepLab снижение точности составило 9% со значения 0,82 до значения 0,74, а значение скорости повысилось на 48% – с 27 до 40 кадров/с.

В выполненных тестах при обучении нейросетевых алгоритмов DeepLab и DL_Xception_4 применялись одни и те же методы аугментации данных, то есть методы расширения вариативности набора данных при обучении нейросетевого алгоритма за счет изменения некоторых параметров [8 - 10], таких, как например, увеличение или уменьшение масштаба изображений, реализованные в исходном коде программной библиотеки DeepLab. В рамках тестирования для данного преобразования в исходном коде программной библиотеки DeepLab по умолчанию выставлены следующие значения:

- «min_scale_factor» = 0,5 (минимальное значение коэффициента масштабирования);
- «max_scale_factor» = 2 (максимальное значение коэффициента масштабирования);
- «scale_factor_step_size» = 0,25 (значение шага для изменения коэффициента масштабирования).

С целью проверки возможности увеличения точности нейросетевых алгоритмов DeepLab и DL_Xception_4 за счет расширения набора применяемых методов при аугментации изображений, для каждого изображения из набора данных, представленного выше, а также для графических файлов с разметкой было выполнено преобразование в виде поворота изображения относительно центральной точки изображения на три значения угла: 90° , 180° и 270° (Рис. 7).

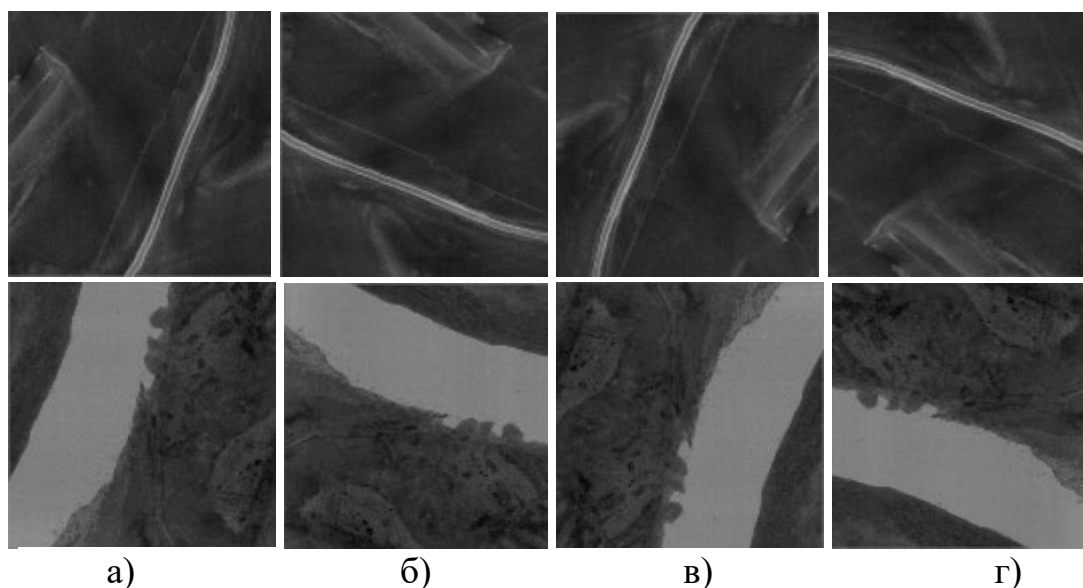


Рис. 7. - Пример выполнения преобразований из набора данных для двух различных изображений, где а) - примеры исходных изображений; б, в, г) - изображения, полученные из исходных в результате поворота на 90° , 180° и 270° относительно центральной точки соответственно

Таким образом, получен расширенный набор данных, состоящий из 4400 изображений. Кроме того, в исходном коде программной библиотеки DeepLab расширен диапазон значений коэффициента изменения масштаба, а также уменьшен шаг изменения коэффициента масштабирования:

- «min_scale_factor» = 0,1;
- «max_scale_factor» = 2,5;
- «scale_factor_step_size» = 0,1.

Далее было выполнено обучение двух нейросетевых алгоритмов DeepLab и DL_Xception_4. Параметры для обучения каждого из указанных

алгоритмов были одинаковы, а именно, использована предобученная модель [4], обучение выполнялось на протяжении 480000 итераций, параметр режима обучения «FINE_TUNE_BATCH_NORM» установлен в значение «false», параметры нейросетевой модели «Atrous rate»: AR = {6; 12; 18} и «Output stride»: OS = 16. Полученные результаты для нейросетевых алгоритмов DeepLab и DL_Xception_4 представлены в таблице № 4 и таблице № 5 соответственно.

Таблица № 4

Оценка точности нейросетевого алгоритма DeepLab (4400 изображений).

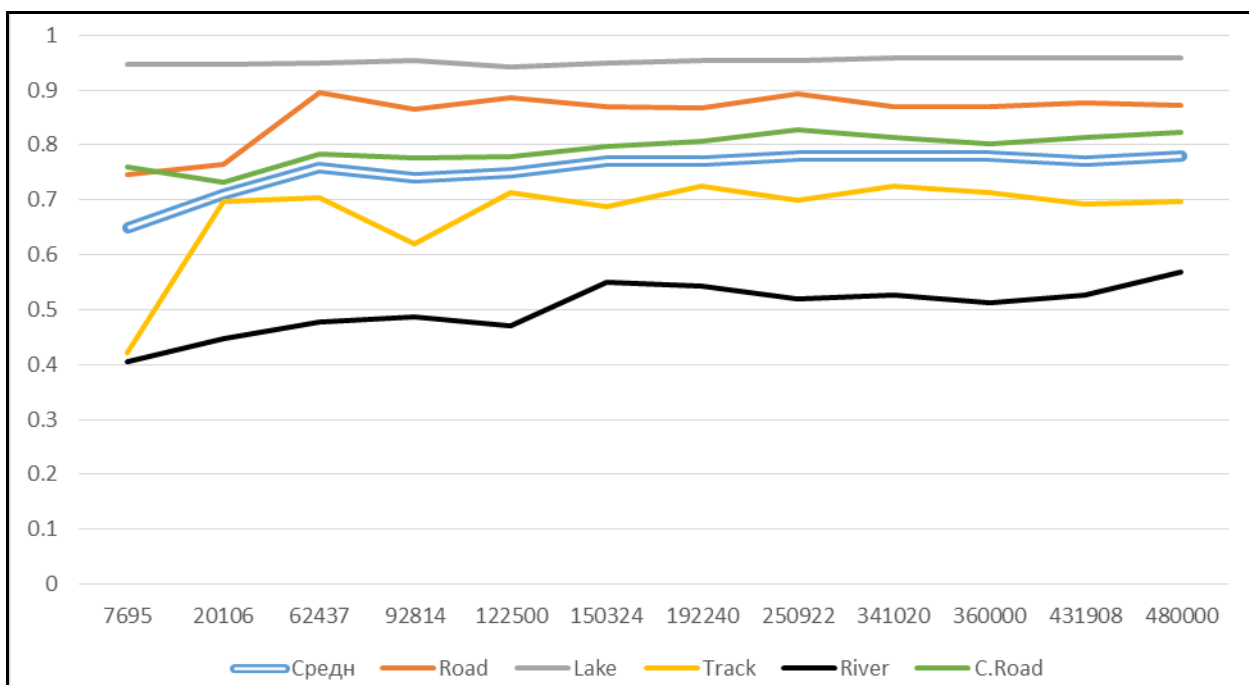
№ итерации	Точность, P, отн. ед.					
	Средн.	Road	Lake	Track	River	C.Road
7695	0,65	0,74	0,94	0,42	0,40	0,76
62437	0,76	0,89	0,95	0,70	0,47	0,78
92814	0,74	0,86	0,95	0,61	0,48	0,77
122500	0,75	0,88	0,94	0,71	0,46	0,77
150324	0,77	0,86	0,94	0,68	0,55	0,79
192240	0,77	0,86	0,95	0,72	0,54	0,80
250922	0,78	0,89	0,95	0,69	0,51	0,82
341020	0,78	0,86	0,95	0,72	0,52	0,81
360000	0,78	0,86	0,95	0,71	0,51	0,80
431908	0,77	0,87	0,95	0,69	0,52	0,81
480000	0,78	0,87	0,95	0,69	0,56	0,82

Таблица № 5

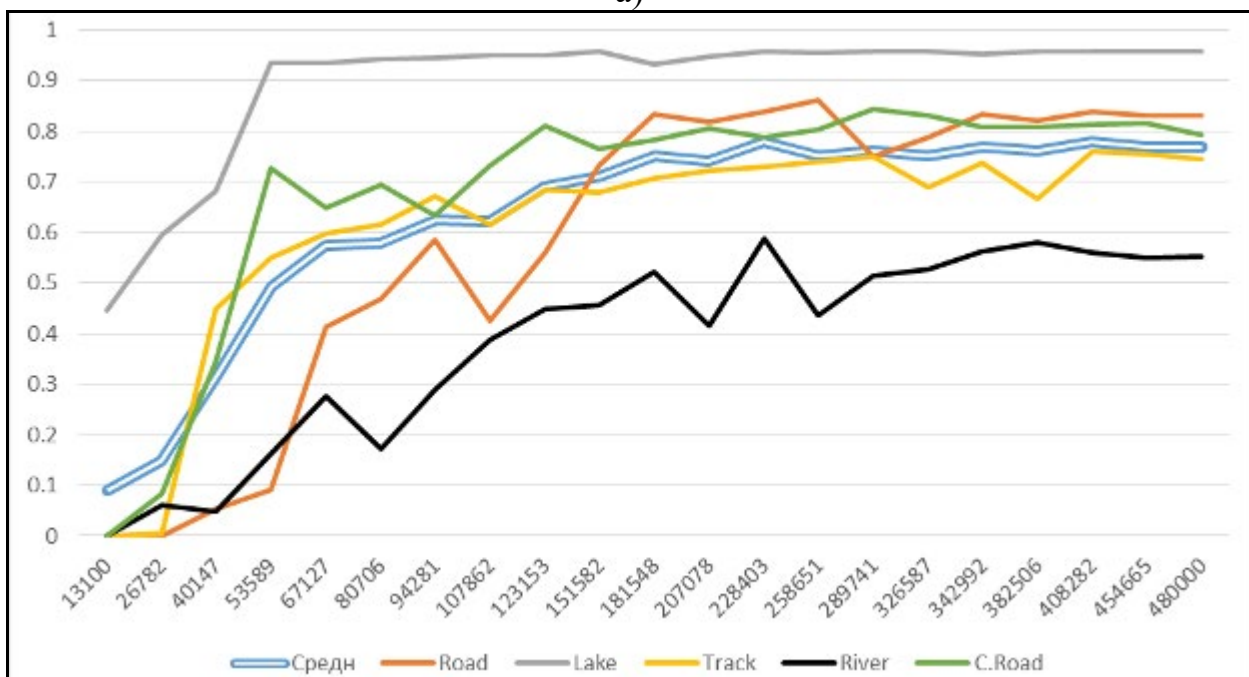
Оценка точности нейросетевого алгоритма DL_Xception_4 (4400 изображений).

№ итерации	Точность, P, отн. ед.					
	Средн.	Road	Lake	Track	River	C.Road
13100	0,08	0	0,44	0	0	0
40147	0,31	0,05	0,68	0,44	0,04	0,34
123153	0,69	0,56	0,94	0,68	0,44	0,81
151582	0,71	0,73	0,95	0,67	0,45	0,76
181548	0,75	0,83	0,93	0,70	0,52	0,78
207078	0,74	0,81	0,94	0,72	0,41	0,80
258651	0,75	0,86	0,95	0,74	0,43	0,80
342992	0,77	0,83	0,95	0,73	0,56	0,80
408282	0,78	0,83	0,95	0,75	0,56	0,81
454665	0,77	0,83	0,95	0,75	0,55	0,81
480000	0,77	0,83	0,95	0,74	0,55	0,79

Графики изменения точности семантической сегментации в процессе обучения для нейросетевых алгоритмов DeepLab и DL_Xception_4 представлены на Рис. 8 а и Рис. 8 б соответственно.



а)

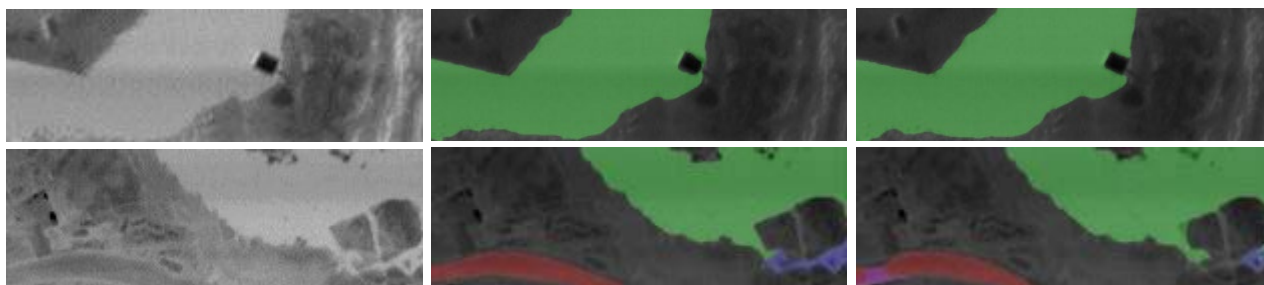


б)

Рис. 8. - Графики изменения точности семантической сегментации в процессе обучения для нейросетевых алгоритмов: а) - DeepLab, б) - DL_Xception_4.

По результатам выполненных тестов на расширенном наборе данных и с увеличением диапазона значений коэффициента изменения масштаба изображений при обучении максимальная точность модифицированного нейросетевого алгоритма DL_Xception_4 повысилась на 5% по сравнению с обучением на исходном наборе данных со значения 0,74 до значения 0,78, а для исходного нейросетевого алгоритма DeepLab максимальное значение точности стало ниже и изменилось со значения 0,82 до значения 0,78.

При визуальной оценке результатов выполнения семантической сегментации (Рис. 9) лучшими по значению точности обученными нейросетевыми моделями исходного алгоритма DeepLab и разработанного алгоритма DL_Xception_4 на объектах с четкими границами разница в качестве не заметна, лишь на объектах с нечеткими границами заметна разница, выраженная в менее плавных переходах и небольших разрывах контуров объектов.



а) б) в)
Рис. 9 а) - исходные изображения, б) - семантическая сегментация алгоритмом DeepLab, в) - семантическая сегментация разработанным методом DL_Xception_4

Выводы

Модификация нейросетевого алгоритма семантической сегментации DeepLabv3+ за счет уменьшения количества параметров нейросетевой модели в кодере на основе нейросетевого алгоритма Xception позволила увеличить скорость обработки графической информации модифицированным

нейросетевым алгоритмом DL_Xception_4 на 48% с 27 до 40 кадров в секунду.

Предложенная методика обучения модифицированного нейросетевого алгоритма DL_Xception_4, заключающаяся в увеличении количества изображений в наборе данных для обучения путем выполнения разворотов изображений вокруг центральной точки изображений, а также заключающаяся в увеличении диапазона значений коэффициента изменения масштаба изображений при обучении, позволила получить точность модифицированного нейросетевого алгоритма DL_Xception_4 до значения 0,78, что на 5% превышает точность для данного алгоритма по сравнению с обучением на исходном наборе данных и что на 5% ниже точности исходного нейросетевого алгоритма DeepLab, значение которой составляет 0,82.

Таким образом, разработанный модифицированный нейросетевой алгоритм DL_Xception_4 со скоростью обработки графической информации 40 кадров в секунду и точностью 0,78 может быть применен для дешифрирования графической информации в составе высокопроизводительных съемочных систем в режиме реального времени непосредственно в процессе выполнения съемочных работ при решении задач в лесном хозяйстве, задач экологического мониторинга, мониторинга стихийных бедствий.

Литература

1. Ямашкин С.А., Ямашкина Е.О., Ямашкин А.А. Репозиторий нейросетевых моделей для анализа пространственных данных // Инженерный вестник Дона, 2022, №12. URL: ivdon.ru/magazine/archive/n12y2022/8082/
2. Chen L., Yukun Z., Papandreou G., Schroff F., Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation // URL: arXiv.org/abs/1802.02611.

3. Peng H., Xiang S., Chen M., Li H., Su Q. DCN-DEEPLABV3+: Novel road segmentation algorithm based on improved DEEPLABV3+// IEEE Access. 2024. T. 12. pp. 87397-87406.
4. deeplabv3_pascal_train_aug_2018_01_04.tar.gz // URL: download.tensorflow.org/models/deeplabv3_pascal_train_aug_2018_01_04.tar.gz.
5. TensorFlow. An end-to-end platform for machine learning // URL: tensorflow.org.
6. DeepLab: Deep Labelling for Semantic Image Segmentation // URL: github.com/tensorflow/models/tree/master/research/deeplab.
7. Everingham M., Ali Eslami S. M., Gool L. V., Williams C. K. I., Winn J., Zisserman A. The Pascal Visual Object Classes Challenge – a Retrospective // International Journal of Computer Vision manuscript No. 111, 98-136 (2015). Doi: 10.1007/s11263-014-07335.
8. Ложкин И.А., Дунаев М.Е., Зайцев К.С., Гармаш А.А. Аугментация наборов изображений для обучения нейронных сетей при решении задач семантической сегментации // International Journal of Open Information Technologies, 2023, №1. URL: cyberleninka.ru/article/n/augmentatsia-naborov-izobrazheniy-dlya-obucheniya-neyronnyh-setey-pri-reshenii-zadach-semanticheskoy-segmentatsii (дата обращения: 28.06.2024).
9. Евдокимова Т.С. Влияние методов расширения наборов данных на качество обучения нейросетевых моделей. Адаптивный подход расширения наборов данных // Инженерный вестник Дона, 2024, №7. URL: ivdon.ru/magazine/archive/n7y2024/9439/
10. Сухаренко Д.В., Шрайнер Б.А. Обучение теме разбиения данных для машинного обучения в целях борьбы с переобучением // Всероссийская научно-практическая конференция «Актуальные проблемы технологического

образования: мастерство, творчество, инновации». Новосибирск, 2022. С. 116-121.

References

1. Yamashkin S.A., Yamashkina E.O., Yamashkin A.A. Inzhenernyj vestnik Dona, 2022, №12. URL: ivdon.ru/magazine/archive/n12y2022/8082/.
2. Chen L., Yukun Z., Papandreou G., Schroff F., Adam H. URL: [arXiv.org/abs/1802.02611](https://arxiv.org/abs/1802.02611).
3. Peng H., Xiang S., Chen M., Li H., Su Q. IEEE Access. 2024. V. 12. P. 87397-87406.
4. deeplabv3_pascal_train_aug_2018_01_04.tar.gz. URL: download.tensorflow.org/models/deeplabv3_pascal_train_aug_2018_01_04.tar.gz.
5. TensorFlow. An end-to-end platform for machine learning. URL: tensorflow.org.
6. DeepLab: Deep Labelling for Semantic Image Segmentation. URL: github.com/tensorflow/models/tree/master/research/deeplab.
7. Everingham M., Ali Eslami S. M., Gool L. V., Williams C. K. I., Winn J., Zisserman A. International Journal of Computer Vision manuscript No. 111, 98-136 (2015). doi:10.1007/s11263-014-07335.
8. Lozhkin I.A., Dunaev M.E., Zaytsev K.S., Garmash A.A. International Journal of Open Information Technologies, 2023, № 1. URL: cyberleninka.ru/article/n/augmentatsia-naborov-izobrazheniy-dlya-obucheniya-neyronnyh-setey-pri-reshenii-zadach-semanticheskoy-segmentatsii, (access date: 28.06.2024).
9. Evdokimova T.S. Inzhenernyj vestnik Dona, 2024, № 7. URL: ivdon.ru/magazine/archive/n7y2024/9439/.
10. Sukharenko D.V., Shrayner B.A. Vserossiyskaya nauchno-prakticheskaya konferentsiya "Aktual'nye problemy tekhnologicheskogo



obrazovaniya: masterstvo, tvorchestvo, innovatsii.”. Novosibirsk, 2022, pp. 116-121.

Дата поступления: 7.01.2025

Дата публикации: 25.02.2025