

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ КЛАССИФИКАЦИИ БИОЛОГИЧЕСКИХ ОБЪЕКТОВ ПО ПОДВОДНЫМ КАМЕРАМ МЭС ОСТРОВА ПОПОВА

Д.Д. Каплуненко, С.С. Зотов, А.Е. Суботэ, В.К. Фищенко

В работе рассматривается применение предобученной сверточной нейронной сети для классификации биологических объектов, попадающих в поле зрения подводной камеры, установленной в бухте Алексеева, на острове Попова, в заливе Петра Великого, в Японском море. Для обзора применимости нейронной сети использовалась модель YOLO5, обученная специально для классификации движущихся подводных обитателей в двух режимах, используемых при обучении (простой и усложненный). Перед началом эксперимента предполагалось, что после обучения модель сможет отличить рыб от креветок, крабов, медуз и морских звезд (всего 6 классов морских биологических объектов). Для обучения модели использовался набор данных, подготовленный университетом г. Аальборг (Дания) и имеющийся в свободном доступе в сети интернет. В дальнейшем обученная на таких данных модель работала с изображениями с камер морской экспедиционной станции (МЭС о-в Попова). Был рассмотрен период, относящийся к осени 2018 г. (сентябрь–октябрь). В процессе работы модели удавалось обнаружить в анализируемых изображениях разные классы морских обитателей с различной достоверностью. В целом проведенные исследования показывают применимость данной модели для проведения подобных работ по классификации, а также необходимость и возможность подготовки собственных наборов данных для обучения таких моделей и применения в научных и хозяйственных целях при работе с региональной спецификой дальневосточных морей.

Ключевые слова: подводное видеонаблюдение, Японское море, модель для детекции объектов YoloV5.

1. Введение

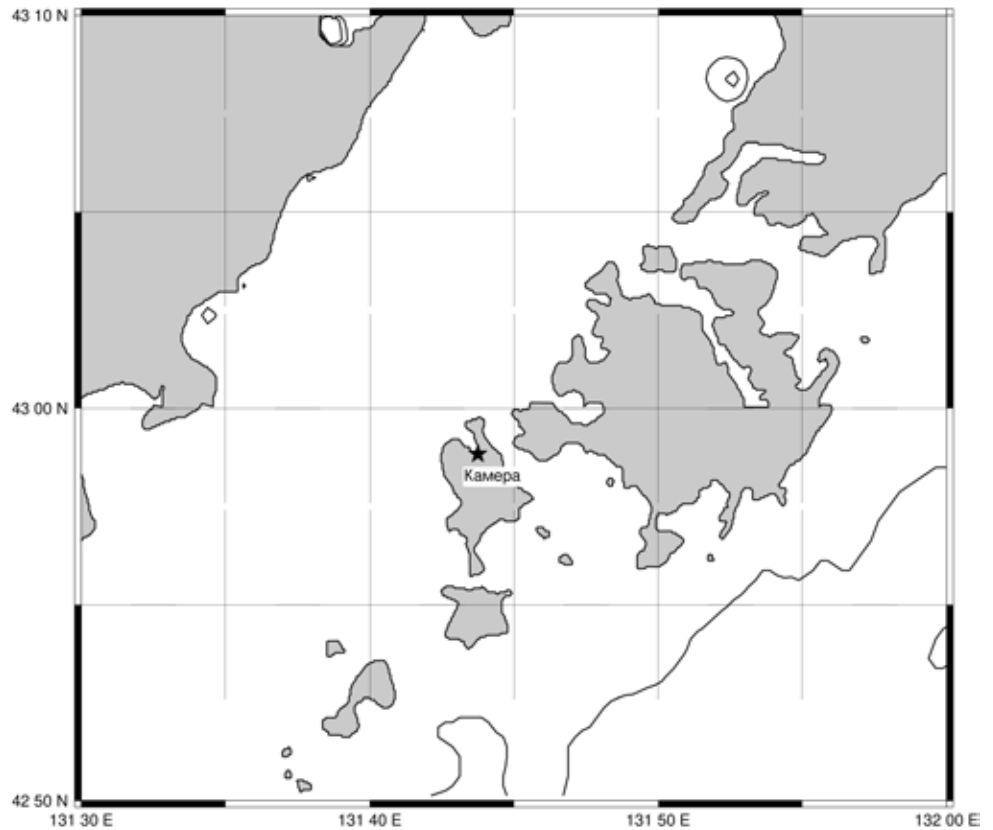
Задача классификации подводных объектов, как живых, так и неживых, является одной из давних и актуальных задач, которые решаются различными способами в последние 30–35 лет с различной степенью успешности [4–6]. При этом накапливаются данные по таким задачам классификации достаточно давно и, в зависимости от возможностей исследователей, к ним применяются различные методы распознавания.

В случае данного исследования была рассмотрена возможность классификации подводных объектов в условиях переменной видимости при наличии минимальной предварительной информации об объекте исследования. В данном случае использовались биологические объекты, появляющиеся в поле зрения подводных камер. Использованные камеры были

установлены в районе бухты Алексеева, на острове Попова, который находится в заливе Петра Великого, юго-западнее острова Русского и г. Владивостока (рис. 1). Характеристики камер детально описаны в публикациях [1, 2]

Также в рамках упомянутых работ [1, 2] рассматривались различные возможности подсчета рыб при помощи специально разрабатываемого ТОИ ДВО РАН программного обеспечения для этих целей (QAVIS и другие). Отметим, что в последнее время получил распространение иной подход к анализу данных, когда для их обработки используются специально подготовленные экспертные системы, предобученные под определенный вид анализа данных. В данном случае использованы нейронные сверточные сети, которые получили широкое распространения при решении задач идентификации объектов при их

Рис. 1. Расположение подводной камеры наблюдения относительно залива Петра Великого и прилегающих островов



минимальном предварительном описании [5, 6]. Для таких случаев применения была поставлена цель изучить применимость популярных, предлагаемых сейчас для использования таких моделей и наборов данных для их обучения [3, 4, 7], а также исследовать возможность их применения в случае данных, получаемых ТОИ ДВО РАН на протяжении последних лет на морских экспедиционных станциях. Таким образом, в данной статье во втором разделе (данные и методы) рассматриваются в качестве методов – класс моделей для детекции объектов (в данном случае – морских обитателей) и в качестве данных – тестовый набор данных для проверки модели и набор данных с подводных камер, подготовленный для анализа. В третьем разделе рассматриваются полученные нами результаты на работе с данной моделью и в четвертом делаются выводы о перспективности дальнейшего применения моделей такого класса в целом для подводных исследований.

2. Данные и методы

Использованная нами в эксперименте модель YOLOv5 является обновленной версией модели YOLO с модифицированной архитектурой на pytorch, предназначенной для распознавания объектов в ре-

альном времени. При работе с потоковым видео для модели YOLOv5 возможно предсказание принадлежности объекта к классу со скоростью до 140 кадров в секунду. Согласно комментариям авторов модели, YOLOv4 и YOLOv5 используют идентичные архитектуры (рис. 2), но улучшение новой модели основано на оптимизации программного кода и использовании более совершенных методов библиотеки PyTorch (https://www.theseus.fi/bitstream/handle/10024/452552/Do_Thuan.pdf стр. 36). Согласно этой архитектуре выполняется предсказание координат определенного количества маркируемых объектов (классы которых определены заранее) и вероятности того, что именно данный объект относится к данному классу. Согласно схеме на рис. 2 это происходит посредством разделения объектов, получаемых со входа детектора (Input) в его одноступенчатой основе (Backbone) с последующим масштабированием и фильтрацией (Neck – часть детектора модели, формирующая детализированные особенности, получаемые из одноступенчатой основы, Backbone) с выходом на вероятностную плотность предсказания (Dense Prediction), как изложено в [3].

Вообще, аббревиатура YOLO (You Only Look Once – вы глядите только однажды) подразумевает семейство моделей, которые в последнее время стали популярны благодаря легковесности (малый объ-

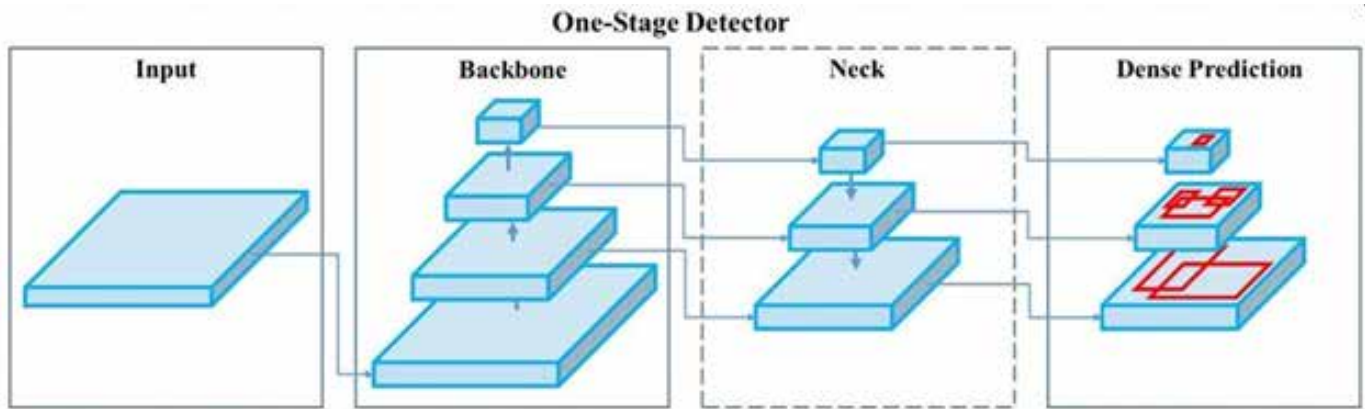


Рис. 2. Схема работы модели YOLOv5 по архитектуре One-Stage detector

ем кода для работы) и качеству предсказаний. Это позволяет использовать YOLO для задач распознавания объектов в реальном времени и на портативных устройствах, например таких, какие используются в проектах IoT (Internet of Things, Интернет вещей). Предыдущие версии, YOLOv2 и YOLOv3 (следует отметить, что в этой работе в дальнейшем используется набор данных Brackish Dataset [4, 8], опробованный на модели YOLOv3 и успешно используемый для обучения YOLOv5), последовательно улучшали качество предсказаний. В апреле 2020 г. была представлена четвертая версия модели YOLOv4. Улучшение качества распознавания проводилось за счет изменений в аугментации данных и архитектуре модели.

В пятой версии модели YOLOv5 в релиз включается 5 моделей разных размеров YOLOv5n, YOLOv5s (маленькая, использована в данной работе), YOLOv5m (средняя), YOLOv5l (большая, также использована в данной работе), YOLOv5x (самая большая). Схемы данных моделей с количеством па-

раметров приведены на рис. 3. Параметры, соответствующие данным моделям, приведены в таблице 1.

Основное из отличий пятой версии модели от других состоит в том, что она специально писалась под библиотеку глубокого обучения PyTorch. Предыдущие версии данной модели, включая четвертую, основывались на нейронной сети Darknet, написанной на языке C и CUDA (более подробно можно ознакомиться здесь: <https://github.com/elsamso/darknet-python>). Поэтому они проигрывали в производительности и возможности конфигурирования. В табл. 1 показаны скорости выполнения предсказаний на кадр с размером 640 пикселей по стороне и при различных вычислительных ресурсах системы. Столбец Speed V100 b32 соответствует параметрам системы, использованным со случаем нашего эксперимента – расчеты на GPU (Graphic Processing Unit), графическом процессоре Tesla V100 с памятью 32 Гбайт, вычислительный кластер ТОИ ДВО РАН.

Расчетное время обучения модели для варианта релиза YOLOv5s в нашем случае составило 3.5 часа

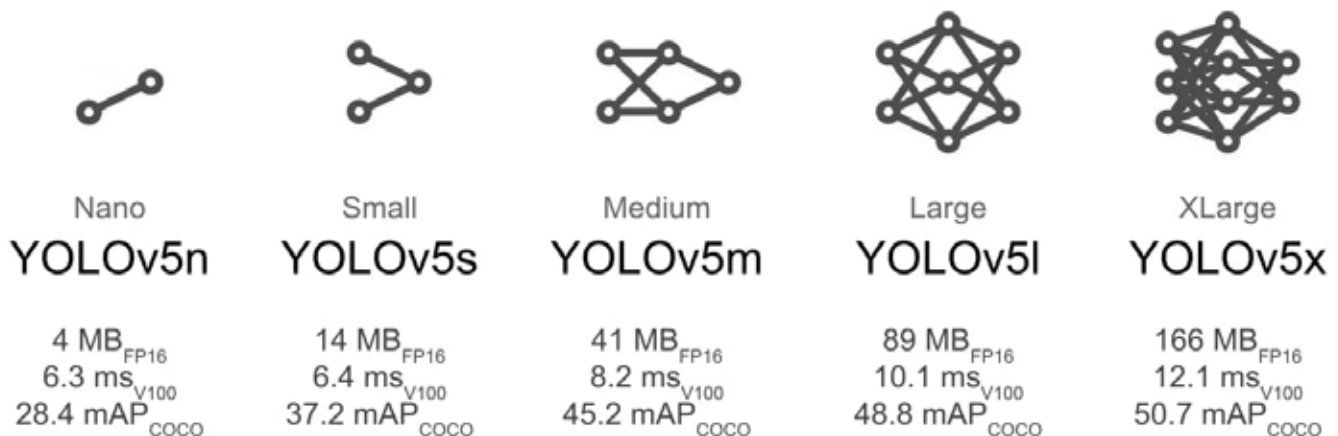


Рис. 3. Разделение моделей по количеству обрабатываемых параметров (FP16, мегабайт), скоростям вычисления (миллисекунды на графическом процессоре Tesla V100) и средней вычислительной точности по обучающей модели COCO (mAP – mean average precision)

Таблица 1. Основные отличия вариантов модели распознавания YOLOv5

Model	size (pixels)	mAP ^{val} _{0.5:0.95}	mAP ^{val} _{0.5}	Speed CPU b1 (ms)	Speed V100 b1 (ms)	Speed V100 b32 (ms)	params (M)	FLOPs @640 (B)
YOLOv5n	640	28.4	46.0	45	6.3	0.6	1.9	4.5
YOLOv5s	640	37.2	56.0	98	6.4	0.9	7.2	16.5
YOLOv5m	640	45.2	63.9	224	8.2	1.7	21.2	49.0
YOLOv5l	640	48.8	67.2	430	10.1	2.7	46.5	109.1
YOLOv5x	640	50.7	68.9	766	12.1	4.8	86.7	205.7

для 150 эпох (итеративных циклов для обучения модели со снижающейся ошибкой mAP), а в случае YOLOv5l для 300 эпох – 16.1 часа.

Обучение выбранных моделей производилось на Brackish Dataset, находящемся в свободном доступе в Internet, на сайтах, связанных с проектом YOLO. Это первый общедоступный европейский набор данных подводных изображений с описательными метками аннотаций рыб, крабов и других морских организмов. Набор данных был сформирован в Лимфьорде, проливе с соленой водой, проходящем через г. Ольборг в северной части Дании. Установка, использованная авторами Brackish Dataset [4], отчасти воссоздает условия, аналогичные тем, которые имеются в бухте Алексева на острове Попова (рис. 1). Камера,

Таблица 2. Основные классы набора Brackish Dataset для модели распознавания YOLOv5, использованные в эксперименте для подводных камер ТОИ ДВО РАН

№ п/п	Класс	Количество изображений	Количество меток класса
1	Рыбы	1467	322
2	Малые рыбы	1467	1148
3	Крабы	1467	1189
4	Креветки	1467	76
5	Морские звезды	1467	791
6	Медузы	1467	55
7	Всего кадров	1467	3581

выполнявшая запись набора данных, находилась на глубине 9 м от поверхности при постоянном освещении, что объясняет поведение морских обитателей, попадающих в радиус действия светодиодного прожектора. Суммарные данные, полученные при обучении модели, приведены в табл. 2. Набор данных Brackish Dataset [4], для удобства применения, предлагается авторами в двух вариантах: набор изображений с разрешением 960×540, 14674 изображения, а также 1920×1080 с таким же количеством изображений [8]. В эксперименте был использован набор изображений с разрешением 960×540.



Рис. 4. Распознавание данных Brackish Dataset на тестовом наборе данных для обучения модели

Перед началом работы с изображениями с подводных камер острова Попова было выполнено тестовое распознавание данных для набора данных, используемого для обучения. Результат приводится на рис. 4.

Для работы с моделью был подготовлен набор данных на основе наблюдений с подводных камер острова Попова, из расположения, указанного на рис. 1. Согласно описанию модель YOLOv5 может выполнять распознавание объектов как для отдельных изображений, так и для потокового видео, включая видеокamerу и видеофайлы различного формата. Поскольку обучение было выполнено на изображениях, то для сравнимости результатов был подготовлен набор данных, состоящих из изображений. Для этого видео с подводных камер разрешением 1280×720 p раскладывалось на опорные кадры (примерно 1 кадр из 25), затем размер картинки приводился к разрешению 640×384 – оптимальное разрешение кадра, близкое к разрешению набора кадров для обучения. Таких кадров было отобрано 22552 и затем выполнен анализ изображений на основе предобученных моделей YOLOv5s (малая) и YOLOv5l (большая). Более подробно полученные результаты рассматриваются в разделе 3. Здесь же отметим, что выделенные классы в модели по набору для обучения могут давать сбой при выполнении процедур распознавания. Типичные результаты, получаемые при распознавании данных с камер острова Попова, приводятся на рис. 5.

3. Результаты

На основе тестов модели YOLOv5s и YOLOv5l было выяснено, что после обучения на наборе данных Brackish Dataset детектирование производится правильно и может быть применено для детекции объектов с подводных камер о-ва Попова. При помощи упомянутых моделей нами был проанализирован набор изображений из опорных кадров, упомянутый в разделе 2, состоящий из 22552 изображений разрешением 640×384 пикселей. На основе данного анализа нами получены количественные и качественные данные, из которых можно делать вывод о пригодности данной модели для изучения выбранного нами предмета. Качественные выводы, т.е. представления о том, насколько детекция объектов справедлива по отношению к нашим данным, были выполнены на основании анализа данных, представленных в табл. 3. По сути, на нем видно, что в отдельных случаях модель классифицирует правильно выделяемые объекты. Как правило, это такие объекты, которые наиболее часто попадались при обучении модели и

указаны в табл. 2. При этом не исключены ошибки, когда один класс, например рыбы, относится к классу крабов. Также очевидно, что европейский набор данных не покрывает все виды рыб, которые встречаются в дальневосточных морях. Например, в табл. 3 видно, что относимый моделью к классу крабов объект № 6 на самом деле является морским ежом.

Кроме данных качественных, разобранных в табл. 3, также, с нашей точки зрения, представляют интерес и количественные данные, из которых видно, насколько меняется процент определенных морских обитателей в зависимости от вида используемой модели. Такие данные нами обработаны по выходным данным модели и сведены в табл. 4 этого раздела.

Как видно из таблицы, с использованием более сложной модели, YOLOv5L, уменьшается количество ложных срабатываний, что видно из строки, где был произведен подсчет пустых кадров, где какие-либо морские обитатели полностью отсутствовали. Остальные данные количественно перераспределяются, однако примерно соответствуют распределению модельных данных, полученных при обучении на Brackish Dataset.

Также следует отметить, что при анализе выходных данных модели не было выявлено ни одной морской звезды, что говорит о различии в модельных весах, необходимых для определения данного класса. Иными словами, морские звезды дальневосточных окраинных морей требуют отдельной классификации, которая отличается от выполненной в Brackish Dataset.

Таким образом, выполненный нами анализ работы модели YOLOv5 позволяет сделать некоторые выводы о применимости данной модели для дальневосточных окраинных морей и перспективах ее применения для классификаций подводных живых и неживых объектов, наблюдаемых стационарными и мобильными подводными камерами.

4. Заключение

В результате экспериментов с применением вычислительного кластера ТОИ ДВО РАН, оснащенного графическим процессором GPU Tesla V100 32 GB, нами было выполнено обучение модели YOLOv5, применяемой для быстрой классификации различных движущихся объектов живой и неживой природы. Основной целью эксперимента было изучение возможной применимости модели для классификации подводных живых объектов, которые динамически и самостоятельно перемещаются в поле зрения подводных видеокamer стационарной установки.

Таблица 3. Результаты качественной детекции моделью YOLOv5s морских животных объектов по подводным камерам о-ва Попова

















№ п/п	Класс	Исходный кадр	Детекция	Результат
1.	Рыбы			Истина, однако есть ложные срабатывания
2.	Рыбы			Истина, однако есть ложные срабатывания
3.	Малые рыбы			Истина
4.	Крабы			Истина
5.	Крабы			Ложь
6.	Крабы			Ложь, объект вне классификации
7.	Креветки			Ложь
8.	Медузы			Возможно истина, но при высокой мутности

Таблица 4. Количественные данные об объектах, идентифицированных моделью YOLOv5 в процессе распознавания данных с видеокamer о-ва Попова

№ п/п	Класс	Количество определений (модель YOLOv5s, малая)	Количество определений (модель YOLOv5l, большая)
1	Рыбы	1525	1694
2	Малые рыбы	120254	30699
3	Крабы	2375	1741
4	Креветки	2644	290
5	Морские звезды	Не было	Не было
6	Медузы	58	146
7	Пустой кадр (никого нет)	1	1860
8	Всего кадров	22552	22552

В результате установлено, что при наличии соответствующего набора данных для обучения модели могут решаться задачи по статистической классификации подводных живых объектов, позволяющие выявлять количественные показатели биоресурсов. Точность данного анализа зависит от используемого набора данных. Имеющиеся сейчас в открытом доступе данные о морских обитателях в европейской части континента недостаточно полны для точных и статистически значимых оценок местных подводных обитателей.

Поэтому представляется перспективным создание собственного набора данных для классифика-

ции морских обитателей дальневосточных морей для использования его как со стационарными, так и перемещающимися камерами подводного наблюдения.

Также следует отметить, что при всех недостатках европейского набора данных Brackish Dataset представляется разумным использовать его в качестве основы для собственного набора данных, с дообучением, так как в целом он показывает свою применимость, выполняя распознавание классов, заложенных в него с достаточно высокой точностью даже относительно другого набора данных, полученного в иных условиях.

ЛИТЕРАТУРА

1. Фищенко В.К., Зимин П.С., Зацерковный А.В., Гончарова А.А., Суботэ А.Е., Голик А.В. Стационарные системы подводного видеонаблюдения: возможности применения для мониторинга биоты прибрежных акваторий залива Петра Великого (Японское море) // Вестн. ДВО РАН. 2018. № 1. С. 149–160.
2. Фищенко В.К., Зимин П.С., Зацерковный А.В., Суботэ А.Е., Голик А.В., Гончарова А.А. Системы стационарного подводного видеонаблюдения прибрежных акваторий // Подводные исследования и робототехника. 2020. № 1 (31). С. 60–71. DOI: 10.37102/24094609.2020.31.1.008
3. Bochkovskiy A., Wang Chien-Yao, Liao Hong-Yuan Mark. YOLOv4: Optimal Speed and Accuracy of Object Detection // arXiv:2004.10934v1 [cs.CV] 23 Apr 2020
4. Pedersen M., Haurum J. B., Gade R., Moeslund T. B., & Madsen N. Detection of Marine Animals in a New Underwater Dataset with Varying Visibility // IEEE Conference on Computer Vision and Pattern Recognition Workshops IEEE. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) – http://openaccess.thecvf.com/content_CVPRW_2019/html/AAMVEM/Pedersen_Detection_of_Marine_Animals_in_a_New_Underwater_Dataset_with_CVPRW_2019_paper.html
5. Thum G. W., Tang S. H., Ahmad S. A., Alrifay M. Toward a Highly Accurate Classification of Underwater Cable Images via Deep Convolutional Neural Network – <http://dx.doi.org/10.3390/jmse8110924>
6. Ricolfe-Viala C., Blanes C. Improving Robot Perception Skills Using a Fast Image-Labeling Method with Minimal Human Intervention – <https://doi.org/10.3390/app12031557>
7. Jocher Glenn. Train Custom Data for YOLOv5 – <https://github.com/ultralytics/yolov5#pretrained-checkpoints> (дата обращения: 1.02.2022).
8. Brackish Underwater Dataset – <https://public.roboflow.com/object-detection/brackish-underwater> (дата обращения: 1.02.2022).

Об авторах

КАПЛУНЕНКО Дмитрий Дмитриевич, к.г.н., старший научный сотрудник лаборатории физической океанологии Федерального государственного бюджетного учреждения науки Тихоокеанский океанологический институт им. В.И. Ильичева Дальневосточного отделения Российской академии наук,
E-mail: dimkap@poi.dvo.ru
Тел. +7 (423) 231-3087, **факс** +7 (423) 231-2573,
Адрес: 690041, г. Владивосток, ул. Балтийская, 43

ЗОТОВ Сергей Сергеевич, старший преподаватель, руководитель образовательной программы ТД33 Академии цифровой трансформации Школы математики и компьютерных технологий Дальневосточного федерального университета,
E-mail: zotov.ss@dvfu.ru
Тел. служ. (423) 2362424 (доб. 2281),
Факс +7 (423) 243-2315
Адрес: 690922, Приморский край, о. Русский, поселок Аякс, 10, кампус ДВФУ, ИМиКТ

СУБОТЭ Алексей Евгеньевич, начальник отдела эксплуатации, развития и защиты информационно-телекоммуникационных ресурсов
 Федеральное государственное бюджетное учреждение науки Тихоокеанский океанологический институт им. В.И. Ильичева Дальневосточного отделения Российской академии наук
E-mail: subote.ae@poi.dvo.ru
Тел. +7 (423) 231-2851, **факс** +7 (423) 231-2573
Адрес: 690041, г. Владивосток, ул. Балтийская, д. 43

ФИЩЕНКО Виталий Константинович, к.т.н., заведующий лабораторией анализа океанологической информации
 Федеральное государственное бюджетное учреждение науки Тихоокеанский океанологический институт им. В.И. Ильичева Дальневосточного отделения Российской академии наук
E-mail: fischenko@poi.dvo.ru
Тел. +7 (423) 231-1631, **факс** (423) 2312573
Адрес: 690041, г. Владивосток, ул. Балтийская, 43