

Методологические подходы и алгоритмы распознавания и подсчета животных на аэрофотоснимках

В. В. Михайлов^а, доктор техн. наук, профессор, ведущий научный сотрудник, orcid.org/0000-0001-6973-3526, mwwcari@gmail.com

Л. А. Колпащиков^б, доктор биол. наук, начальник научного отдела, orcid.org/0000-0003-4871-3360

В. А. Соболевский^а, аспирант, младший научный сотрудник, orcid.org/0000-0001-7685-4991

Н. В. Соловьев^в, канд. техн. наук, доцент, orcid.org/0000-0001-8851-3713

Г. К. Якушев^в, бакалавр, orcid.org/0000-0003-4689-3425

^аСанкт-Петербургский федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ

^бОбъединенная дирекция заповедников Таймыра, Талнахская ул., 22, подъезд 2, Норильск, 663000, РФ

^вСанкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

Введение: сложность процессов распознавания и подсчета объектов на фотоснимке непосредственно связана с изменчивостью сопутствующих факторов: физическим различием объектов одного класса, наличием изображений, сходных с распознаваемыми объектами, неоднородным фоном, изменением условий съемки и положением объектов при получении снимка. В этом смысле сложными являются задачи распознавания людей в скоплениях, животных в природной среде, автомобилей с камер видеонаблюдения, объектов строительства и инфраструктуры на аэрофотоснимках и т. п. Задачи имеют свою специфику факторного пространства, но методические подходы к их решению сходны. **Цель:** разработка методологии и программных реализаций, решающих задачу распознавания и подсчета объектов с высокой изменчивостью, на примере распознавания северных оленей в природной среде. **Методы:** сравнение двух подходов распознавания: по признакам на основе бинарной классификации пикселей и по эталонам с использованием сверточных нейронных сетей. **Результаты:** разработаны методологии и программы для попиксельного распознавания с последующей бинаризацией, кластеризацией изображения и подсчетом числа кластеров и распознавания изображений с использованием сверточной нейронной сети архитектуры Mask R-CNN. При этом сеть первоначально обучается распознаванию животных как класса по массиву изображений MS COCO dataset, затем «доучивается» на массиве аэрофотоснимков стад северных оленей. Анализ результатов показывает, что признаковые методы с попиксельным распознаванием дают хорошие результаты на сравнительно простых изображениях (ошибка распознавания 10–15%). Наличие на снимке артефактов, близких по характеристикам к изображениям оленей, приводит к значительному росту ошибки. Сверточная нейронная сеть показала более высокую точность, которая на тестовой выборке составила 82%, причем ложные срабатывания отсутствовали. **Практическая значимость:** создан программный прототип системы распознавания на основе сверточных нейронных сетей с веб-интерфейсом, а сама программа введена в ограниченную эксплуатацию.

Ключевые слова — распознавание, сверточные нейронные сети, аэрофотоснимки, северные олени.

Для цитирования: Михайлов В. В., Колпащиков Л. А., Соболевский В. А., Соловьев Н. В., Якушев Г. К. Методологические подходы и алгоритмы распознавания и подсчета животных на аэрофотоснимках. *Информационно-управляющие системы*, 2021, № 5, с. 20–32. doi:10.31799/1684-8853-2021-5-20-32

For citation: Mikhailov V. V., Kolpaschikov L. A., Sobolevskii V. A., Soloviev N. V., Yakushev G. K. Methodological approaches and algorithms for recognizing and counting animals in aerial photographs. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2021, no. 5, pp. 20–32 (In Russian). doi:10.31799/1684-8853-2021-5-20-32

Введение

Сложность решения задачи автоматического распознавания объектов по их изображению непосредственно зависит от степени изменчивости факторов, сопутствующих распознаванию. Рассмотрим возможный набор факторов при распознавании объектов на фотоснимках.

1. Распознаваемые объекты как фактор распознавания, которые могут существенно различаться по цвету, размерам, форме — признакам, позволяющим отличать объекты разных классов.

2. Артефакты на фотоснимке, по некоторым признакам сходные с объектами распознаваемого класса.

3. Фон на снимке, который может быть однородным или неоднородным по цвету, может существенно отличаться на разных снимках.

4. Условия освещенности. Источники света могут быть естественными или искусственными, объекты могут находиться на солнце или в тени, снимки могут быть сделаны в разное время суток — все это приводит к изменению яркостных и цветовых характеристик объектов и фона.

5. Расположение объектов распознавания при съемке. Объекты могут находиться на разном удалении и под разными углами к камере, изображения объектов на снимках могут перекрывать друг друга, частично закрываться элементами фона.

Сложность решения задачи распознавания зависит от размерности факторного пространства (количества факторов) и от числа различных значений каждого фактора на снимке. В первом приближении сложность можно оценить по количеству факторов с изменчивостью. Наибольшую сложность имеют задачи, в которых все факторы обладают изменчивостью, наименьшую — задачи с фиксированными значениями факторов распознавания. Структуризация исходного множества снимков, на которых происходит обучение системы распознавания, на подмножества в целях фиксации значений каких-либо факторов позволяет упростить решение задачи распознавания на каждом подмножестве данных, но увеличивает объем ручного труда, затрачиваемого на анализ и группирование снимков.

К сложным в указанном выше смысле в полной мере относятся задачи распознавания различного рода природных объектов: людей в скоплениях, животных и птиц в естественной среде, растительных группировок в зональных и интразональных сообществах, аномалий планетарных морфологических структур (например, лунных кратеров) и т. п. В техносфере это, например, распознавание объектов строительства на аэрофотоснимках, автомобилей по снимкам с камер видеонаблюдения, замаскированных военных объектов по снимкам с беспилотников. Методические подходы к решению указанных задач сходны, однако каждая из них обладает своей спецификой факторного пространства и требует индивидуального подхода для учета этих особенностей и выбора конкретного метода распознавания.

Приведенный выше состав факторов соответствует задаче распознавания северных оленей на аэрофотоснимках. Какие-либо ограничения на значения факторов распознавания не накладывались за исключением масштаба снимков, который соответствует принятому при авиаучетных работах.

Выбор северных оленей как объекта распознавания и учета объясняется:

- соответствием в полной мере задаче с высокой изменчивостью факторов распознавания;
- актуальностью как сокращения ручного рутинного труда специалистов по обработке снимков, так и своевременного получения данных для решения задач управления экологическими процессами Крайнего Севера;
- наличием массива исходных данных, содержащего около 5000 архивных аэрофотосним-

ков стад северных оленей, сделанных при проведении авиаучетов.

Используемые в настоящее время методики учета численности диких северных оленей тундровых популяций (таймырской, якутских популяций, оленей Чукотки, а также мигрирующих стад северных оленей — карибу Канады и Аляски) основываются на экологической особенности видов, состоящей в том, что олени в жаркую погоду во время вылета кровососущих насекомых собираются в многотысячные скопления на ограниченной территории в северной части летнего ареала [1, 2]. Стада в скоплениях фотографируются с летательного аппарата, и количество животных в них подсчитывается операторами вручную по каждому снимку. На одном фотоснимке могут присутствовать от нескольких сотен до 1000 и более оленей. Самое большое стадо, которое было зафиксировано на одном снимке и просчитано «по головам», содержало около 35 000 животных (Таймыр, район р. Верхняя Таймыра, 28.07.2017). Полное число снимков стад при проведении авиаучета может достигать тысячи. В результате прямого подсчета учитывается основная часть оленей в популяции. Количество животных, не вошедших в скопления, оценивается путем аппроксимации по территории [3].

Снимки, сделанные во время авиаучетов, разнообразны по типам ландшафтов, освещенности, направлению движения стад относительно камеры. Олени на снимках имеют разную величину и окраску, что связано как с физиологией животных, так и с условиями освещенности и положением стад. В скоплениях изображения оленей могут перекрывать друг друга, в результате часто видны лишь фрагменты тел. Помехи в виде камней, кочек, впадин затрудняют обнаружение животных. Ручная обработка результатов съемок специалистами-биологами занимает около трех месяцев, в то время как для экологически обоснованного управления динамикой численности популяции, рационального использования биологических ресурсов вида и определения норм промыслового изъятия оленей данные желательно иметь через 10–15 дней после окончания авиаучета.

О принципах и подходах к распознаванию природных объектов

Задача распознавания объектов живой природы имеет большое познавательное значение, поскольку ее решение позволяет увидеть природу в ее многообразии, выделить в общей мозаике по внешним признакам конкретные объекты, узнать их видовую принадлежность, найти их научное наименование и далее, при желании, узнать

их свойства, место и роль в экосистеме. С научных позиций процесс распознавания и учета дает возможность выявить видовой состав экосистемы и динамику ее компонент. В природоохранной деятельности подобная информация является основой для принятия решений о мерах по охране или по ограничению видов или отдельных популяций. Для промысловых видов животных регулярные учеты численности необходимы при определении лимитов изъятия, обеспечивающих рациональное использование биологических ресурсов.

Исторически задача распознавания биологических объектов решалась путем создания определителей, позволяющих установить место данного объекта в общей системе биологической классификации, т. е. установить видовой, родовой, семейный и другие уровни принадлежности объекта. Методически задача распознавания решается в определителях с использованием двух принципиально различных подходов. Первый — распознавание по признакам. Разработчиками определителей такого рода являются специалисты в области биологической систематики и классификации. Они определяют признаки, позволяющие объединить объекты на некотором уровне классификации, и признаки, по которым объекты переходят в разные группы на более низкий уровень классификации и различаются уже в них. В результате формируется дерево с ветвлениями по ключевым признакам. Распознавание объекта сводится к движению по дереву от корня (или какого-либо другого уровня классификации) к ветвям с выбором направления в точках ветвления в соответствии со значениями признаков. Второй подход — распознавание «по образцу». Определитель содержит множество эталонных изображений объектов разных видов. Методика распознавания состоит в сопоставлении искомого объекта с эталонами для выбора наиболее «похожего» образца. Пользователь выступает здесь в роли эксперта, оценивающего степень сходства изображений. Соответственно, ошибка распознавания будет зависеть от опыта, знаний эксперта и его персональных особенностей визуального восприятия объектов. Для распознавания на видовом уровне данный подход может быть весьма эффективным, поскольку основан на сопоставлении и сравнении целостных зрительных образов, что является естественной рутинной задачей зрительной системы человека. Во многих случаях определители по признакам дополняются набором изображений-эталонов, а человек при оценке «схожести» мысленно выделяет признаки сходства и различия объектов.

Компьютерные системы распознавания объектов живой природы также строятся с использованием этих двух подходов. Системы распознавания по признакам являются «прозрачными»

относительно состава используемых признаков объекта распознавания, методов, технологической цепочки преобразований изображения и результатов, включая промежуточные. Однако при распознавании объектов с высокой изменчивостью в нестационарной среде возникают сложности в связи с изменением параметров факторов распознавания. Появляется необходимость поиска новых систем признаков со сменой методов обработки изображений. В работе [4] высказывается предположение о принципиальной невозможности полной формализации процессов распознавания объектов в подобных условиях и необходимости перехода от автоматической к интерактивной технологии обработки изображений.

В современных системах распознавания по эталонам в качестве распознающего субъекта широко используются искусственные нейронные сети (ИНС) [5, 6]. В таких системах ИНС обучается на множестве изображений, в которых объекты распознавания (эталоны) заранее помечены. Затем обученная ИНС используется для распознавания объектов данного класса в реальной среде, на фотоснимках и видеокдрах. Логика распознавания объектов инкапсулирована в самой ИНС и не требует разработки в явном виде. С другой стороны, из-за сложности структур ИНС интерпретация результатов их работы зачастую затруднена. А сам процесс генерации и «обучения» таких структур происходит, по большей части, эмпирически. На сегодня не существует устоявшихся методик создания ИНС, которые бы подходили для большинства прикладных задач. Чаще всего имеются лишь общие рекомендации и эмпирический опыт разработчиков в решении конкретного класса задач.

Известные по литературным источникам системы распознавания и подсчета объектов по фотоснимкам ориентированы главным образом на человека как объект наблюдений. В меньшей степени это относится к другим видам живых организмов. Рассмотрим некоторые из таких систем.

В статье [7] представлена программа распознавания животных и их поведения в дикой природе по фотоснимкам с 225 камер-ловушек заповедника Серенгети (Танзания). Причина разработки системы заключалась в том, что из-за ограниченности в людских ресурсах реально используется лишь малая часть той информации, которая хранится на серверах. Программа распознавания разрабатывалась совместно в университетах Вайоминга, Миннесоты, Обернска, Гарварда, Оксфорда и в компании Uber AI. В работе использовался метод глубокого обучения различных архитектур сверточных нейронных сетей (СНС) [8]. Были использованы 1,5 млн изображений, из которых 1,4 млн служили для обучения сети, 0,1 млн — для тестирования. По результатам те-

стирования вероятность обнаружения животных с бинарным ответом «да/нет» составила около 96 %. Вид животных из 48 возможных вариантов распознавался с вероятностью 99 %. Однако подсчет животных по кластерам 1, 2, ..., 10, 11–50, > 50 выполнялся программой с точностью лишь 63 %. Это говорит о сложности фиксации отдельного животного на снимке в условиях возможных наложений изображений животных друг на друга, их частичного перекрытия посторонними предметами и других искажений изображения животных.

Сверточные нейронные сети применялись также для распознавания слонов на космоснимках [9]. Для обучения сети использовались архивные снимки со спутников Worldview-3 и Worldview-4 с разрешением 31 см, сделанные в период с 2014 по 2019 год на территории национального парка Addo Elephant в Южно-Африканской Республике. На площади свыше 1,6 тыс. км обитают более 600 саванных слонов. Парк имеет неоднородный ландшафт: там есть кустарники, невысокие леса, открытые луга, водоемы. В течение дня слоны перемещаются по территории и иногда обливают себя грязью, чтобы защититься от жары. Все это затрудняет их обнаружение.

Разработчики протестировали систему на независимой выборке снимков территорий Кении и сравнили точность ее работы с результатами 51 добровольца, которые идентифицировали слонов вручную. Среди участников эксперимента были специалисты по машинному обучению, ученые, работники национальных парков. Вероятность обнаружения слонов на гетерогенной территории с использованием СНС составила 78 %, люди находили слонов с вероятностью 77 %. Для однородной территории результаты составили соответственно 73 и 80 %.

Разработана и протестирована система автоматического распознавания, сегментации и подсчета морских котиков в рамках онлайн соревнования платформы Kaggle по Data Science [10]. Постановка задачи связана с мониторингом популяции котиков Западных Алеутских островов, который ведется с использованием дронов. Обработка полученных фотоснимков выполняется вручную, что занимает около четырех месяцев каждый год. Подсчет особей выполняется по пяти половозрастным классам. Для разработки автоматической системы был сформирован массив из 948 размеченных фотоснимков и еще 18 641 снимок представлен для тестирования системы. Снимки сделаны с различной высоты и в разном масштабе, имеются помехи, сходные с изображениями животных, между некоторыми классами котиков отсутствуют четкие визуальные различия. Эти причины усложняют решение задачи автоматического распознавания и подсчета котиков

на аэрофотоснимках. Предложенный вариант решения основан на использовании базовой архитектуры СНС Inception-Resnet-V2, которая расширена добавлением слоев average-pooling + FC-слой на 256 нейронов + Dropout + финальный FC-слой на пять нейронов. Рейтинг проекта — 4-е место среди 385 участников. Особенность данной статьи состоит в том, что в ней не только дается описание конкретной системы, но детально рассматривается весь процесс ее разработки с анализом правильных и ошибочных решений при подборе архитектуры и параметров нейронной сети.

В программах Seek iNaturalist распознавания различных видов растений и животных для любителей природы использованы СНС на основе SSD (Singl-Shot Multy Box Detector) [11]. Такая архитектура дает возможность применять миниатюрные и мобильные устройства для фото- и видеоанализа изображений в реальном времени.

Можно сделать несколько замечаний, касающихся построения автоматических систем распознавания животных на основе существующих представлений о решении задачи классификации и имеющихся программных разработок:

- животное в природной среде — это сложный объект для распознавания. Изменчивость изображений на фотоснимках, их наложение друг на друга, наличие помех приводят к ошибкам распознавания, допускаемых как человеком, так и автоматической системой;

- наиболее приспособленным к решению задач распознавания изображений природных объектов в настоящее время является аппарат СНС;

- повысить точность распознавания можно за счет увеличения объема обучающей выборки или путем перехода к интерактивному режиму обработки снимков с участием человека;

- современные архитектуры СНС дают возможность создавать мобильные приложения для работы в полевых условиях непосредственно при проведении учетных работ.

Распознавание животных на аэрофотоснимках по признакам

Исходя из общих принципов построения систем распознавания образов [12], решение задачи распознавания и подсчета животных по признакам укрупненно можно провести в несколько этапов:

- 1) предварительная обработка снимка для повышения различимости объектов;
- 2) бинаризация, т. е. нахождение пикселей, относящихся к изображениям объектов;
- 3) сегментация, т. е. выделение связанных групп пикселей, относящихся к отдельному объекту;
- 4) определение числа объектов.

При разработке системы подсчета оленей следует учесть, что для дальнейшего использования результатов точность подсчета должна находиться в пределах 10–15 %. Таким образом, ошибки в указанных пределах при выделении оленей (пропуск отдельных объектов, их ложное выделение, неточность при подсчете числа оленей в слитных сегментах) вполне допустимы.

На первом этапе производится очистка изображения от помех и шумов. В качестве фильтров могут использоваться различные усредняющие пространственные фильтры [13]. На этом этапе могут быть увеличены контрастность и резкость изображения, изменен масштаб, выполнен переход из одной цветовой модели в другую. Если объекты имеют хорошо различимые резкие границы, то предварительная обработка, как правило, включает и методы выделения границ (фильтры Робертса, Кирша, Превитта, Лапласа [13]).

Аэрофотоснимки выполняются в цвете, т. е. каждый пиксель описывается тремя цветовыми составляющими — R (красный), G (зеленый) и B (синий). Приведение изображения к монохромному (в градациях серого) виду может упростить последующий процесс бинаризации [4].

Процесс бинаризации можно представить как классификацию пикселей снимка, т. е. разделение их на два класса. Важной задачей классификации является выбор признаков, которые должны удовлетворять следующим условиям:

- значения признаков для объектов одного класса должны различаться слабо, а для объектов разных классов — существенно;

- набор признаков должен быть полным, т. е. должен обеспечивать идентификацию объектов любого из классов;

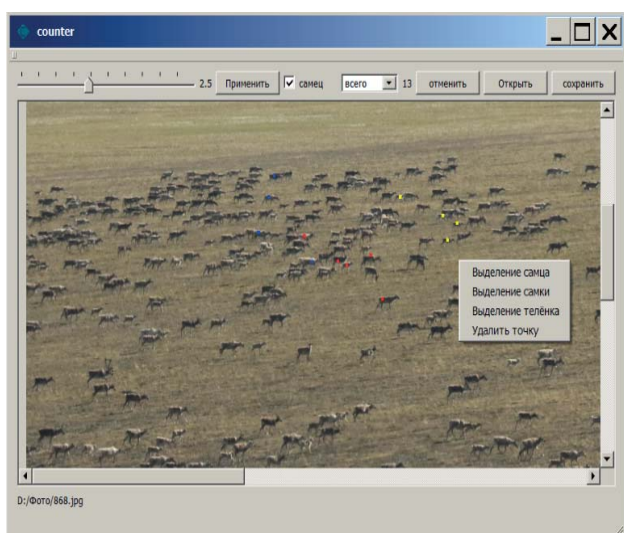
- общее количество признаков должно быть минимальным.

В качестве признаков пикселя можно использовать значения составляющих его цвета, причем как все вместе, так и в различных сочетаниях или по отдельности. Для разделения пространства признаков на два класса необходимо предварительно по обучающей выборке снимков собрать статистические данные о пикселях, относящихся к оленям, в ручном режиме. Для решения этой достаточно трудоемкой задачи была разработана программа [14], позволяющая отображать на дисплее снимок, менять его масштаб, выбирая наиболее подходящий для распознавания объектов, ставить на снимке три разные по цвету метки для самца, самки и теленка, выполнять подсчет выделенных объектов и сохранять результаты (координаты меток и их число) в текстовом файле (рис. 1).

В одной из первых работ по распознаванию и подсчету северных оленей [15] в качестве признака выделения оленей и подстилающей поверхности были приняты различия величин их спектральной яркости. По результатам экспериментов выбрана цветовая пара: зеленый (фон) и голубой (распознаваемый объект), — после чего было составлено неравенство, по которому пиксели объекта отличаются от пикселей фона. Отнесение сегмента к изображению оленя выполнялось по результатам анализа площади и вытянутости сегментированной области. Тестирование показало весьма высокую точность распознавания и подсчета оленей на искусственно сгенерированных «простых» снимках (зеленый фон, отсутствие помех и наложений объектов друг на друга). Ошибка подсчета численности составила около 2 %. Однако при работе с реальными снимками величина ошибки резко возросла, что связано с высокой изменчивостью распознаваемых объектов и условий съемки.

Задачу бинаризации можно представить, как хорошо известную в распознавании образов [12, 13] кластеризацию по k -средним (в данном случае число кластеров $k=2$) в пространстве признаков RGB. Основное преимущество метода — отсутствие необходимости собирать статистические данные перед его применением. Эксперименты на реальных снимках с отсутствием помех показали достаточно хорошие результаты — ошибок около 10 %, но на снимках с неравномерным фоном и наличием помех число ошибок резко выросло. Данный метод наиболее целесообразно использовать для предварительной обработки снимков на этапе сбора статистических данных с последующей корректировкой ошибок в ручном режиме.

Пороговый метод бинаризации пикселей цветного снимка требует установки порогов для отдельных цветовых составляющих с их последующим совместным использованием. Например, для цветовой схемы HSB (H — цветовой тон, S —



■ **Рис. 1.** Интерфейс программы ручной обработки снимков

■ **Fig. 1.** Manual image processing program interface

насыщенность, B — интенсивность) в литературе [16] предлагается для бинаризации использовать только цветовой тон и интенсивность, причем относить к объекту только те пиксели, у которых каждая составляющая меньше соответствующего ей порога, т. е.

$$Z(x, y) = \begin{cases} 1 & H(x, y) < T_H \vee B(x, y) < T_B, \\ 0 & \text{в остальных случаях} \end{cases}, \quad (1)$$

где $Z(x, y)$ — значение яркости пикселя на бинаризованном изображении; $H(x, y)$ — цветовой тон пикселя с координатами (x, y) ; T_H — пороговое значение для цветового тона; $B(x, y)$ — интенсивность пикселя с координатами (x, y) ; T_B — пороговое значение для интенсивности.

Порог T определяется по формуле

$$T = \frac{\sum_x \sum_y I(x, y) \Delta I(x, y)}{\sum_x \sum_y \Delta I(x, y)}, \quad (2)$$

где $I(x, y)$ — значение выбранной для нахождения порога составляющей цветовой модели в пикселе с координатами (x, y) ; $\Delta I(x, y)$ — градиент составляющей цветовой модели:

$$\Delta I(x, y) = \max(\text{abs}(I_x(x, y)), \text{abs}(I_y(x, y))), \quad (3)$$

$$I_x(x, y) = I_x(x + 1, y) - I_x(x - 1, y);$$

$$I_y(x, y) = I_y(x, y + 1) - I_y(x, y - 1). \quad (4)$$

Алгоритм сегментации бинарных изображений хорошо известен [10] и достаточно надежен. В результате его работы каждое связанное множество пикселей получает свой порядковый номер и при необходимости может быть выделено в отдельное изображение для последующего подсчета объектов. Основная проблема, возникающая при подсчете, состоит в том, что на некоторых участках снимка отдельные олени могут сливаться в одно множество сегментированных пикселей. Для решения этой проблемы необходимо проводить анализ распределения сегментов по площади (число пикселей в сегменте), что позволяет оценить среднюю площадь одного объекта и, соответственно, с достаточной точностью найти как число объектов в слитном сегменте, так и общее число объектов на снимке. Сегменты, площадь которых существенно меньше средней площади объекта, целесообразно считать артефактами и не учитывать при подсчете оленей.

Для проведения экспериментов и сравнения результатов применения различных методов распознавания оленей по признакам из имеющихся 100 снимков были сделаны выборки снимков с близкими характеристиками фона и окраса оленей. Для снимков каждой выборки был про-

изведен ручной подсчет оленей и поставлены эксперименты по автоматическому подсчету оленей. Для проведения экспериментов была разработана программа, позволяющая выбирать метод бинаризации, устанавливать необходимые параметры обработки, визуализировать результаты и сохранять количественные показатели для последующей статистической обработки. Сравнение результатов экспериментов с данными ручного подсчета оленей показало, что наиболее близкое совпадение ($\pm 15 \dots 20$ %) дает бинаризация по порогам цветового тона и насыщенности. Наличие на снимке артефактов, близких по яркостным характеристикам, форме и площади к изображениям оленей на данном снимке, приводит к значительному росту ошибки подсчета оленей.

Результаты экспериментов показали, что система распознавания и подсчета оленей на аэрофотоснимках по признакам должна базироваться на детальном анализе и классификации изображений, сборе статистических данных о характеристиках пикселей, относящихся к изображению оленей, определении наиболее информативных признаков и выборе эффективных алгоритмов классификации пикселей и подсчета оленей. Система должна обладать достаточными интеллектуальными способностями для комбинации методов распознавания оленей на предьявляемом снимке.

Распознавание животных на аэрофотоснимках по эталонам с использованием ИНС

СНС как инструмент распознавания изображений

Для решения задачи распознавания объектов живой природы путем обучения по их эталонным изображениям в настоящее время все чаще начинают применять СНС [17]. Данный класс архитектур ИНС является узкоспециализированным инструментом, подходящим прежде всего для работы с изображениями и иными данными, которые возможно представить в матричном виде.

Поскольку в изображениях вся информация хранится в виде двумерных матриц (т. е. в виде пикселей), при работе с ними требуется учитывать значения, поступающие не только от самих нейронов, но и от группы ближайших нейронов. Для этого в сверточных слоях СНС, помимо нейронов, имеется еще один тип элементов, которые применяют определенные линейные операции ко всем входным данным каждого нейрона слоя, — ядро свертки. Ядро свертки представляет из себя сетку, которая «скользит» по изображению (или по сверточному слою предыдущего уровня) и ищет шаблоны и паттерны в данных.

Если будет найдена такая часть изображения, которая совпадет с шаблоном ядра, то оно передаст расчетному нейрону текущего слоя большее положительное значение. Если совпадения не будет, то ядро передаст небольшое значение или нуль.

Из-за того, что ядро свертки применяется к каждой позиции изображения, сверточный слой СНС является крайне эффективным в задачах обработки изображений, поскольку особенности или закономерности в изображениях могут проявляться в любом месте этих изображений. То есть СНС, в отличие от прочих ИНС, способна анализировать контекстно-зависимые данные.

При этом сверточные слои в СНС могут встречаться несколько раз. И тогда каждый последующий слой получает данные от предыдущего сверточного слоя как от входного, поскольку сверточный слой также является двумерным массивом, и вектор его выходных значений может быть представлен в виде изображения. И в таком случае каждый последующий сверточный слой может находить все более сложные шаблоны и паттерны изначального входного массива.

Для поставленной задачи была выбрана архитектура Mask Regions with Convolution Neural Networks (MRCNN) [18] (рис. 2), являющаяся подвидом классической СНС. За счет усложнения архитектура более успешно справляется с задачами семантической и объектной сегментации изображений [19, 20].

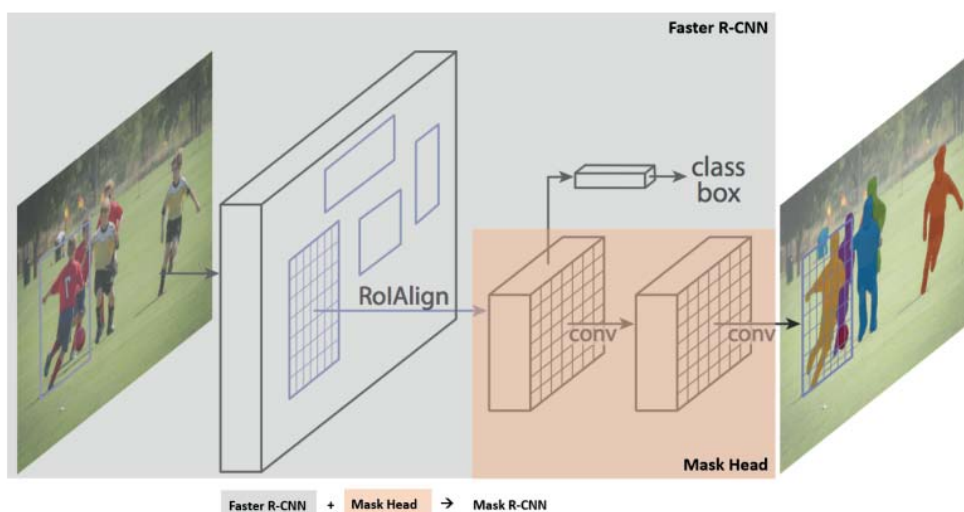
Ключевой особенностью данной архитектуры является совмещение такой архитектуры СНС, как Faster R-CNN (FRCNN), которая отвечает за решение задачи классификации, с модулем Mask Head, который отвечает за решение задачи

сегментации изображения. Результатом работы MRCNN является совмещенный ответ двух описанных составных частей.

Faster R-CNN — это СНС, которая осуществляет поиск на изображении объектов, после чего дополнительно производит классификацию найденного объекта. Результатом работы данной ИНС являются ограничивающие прямоугольники для каждого объекта (т. е. прямоугольная граница, которая ограничивает найденный объект) и метка класса найденного объекта с оценкой достоверности.

Первый этап работы FRCNN состоит из одновременного запуска двух включенных ИНС: магистральной (ResNet, VGG, Inception или аналогичных) и сети регионального позиционирования. Эти сети обрабатывают каждое поступающее на вход FRCNN изображение и на выходе предоставляют трехмерный массив — массив предложенных двумерных областей. Данные двумерные области являются вырезанной частью изображения, на которой магистральная СНС распознала какой-либо объект. Размер данных частей не фиксирован и зависит от размера распознанного объекта.

На втором этапе FRCNN прогнозирует координаты ограничивающих прямоугольников и классы объектов для каждой из предложенных областей, полученных на первом этапе. Каждая предлагаемая область может иметь разный размер, но так как сверточные слои в СНС всегда требуют вектора фиксированного размера для прогнозирования, на данном этапе также производится масштабирование найденных регионов. Размер регионов масштабируется с помощью либо алгоритма RoI, либо метода RoIAlign.



■ Рис. 2. Обобщенная схема MRCNN [18]
 ■ Fig. 2. Generalized MRCNN scheme [18]

MRCNN в свою очередь является расширенной версией FRCNN, дополненной ветвью для прогнозирования масок сегментации для каждой области интереса. На втором этапе работы MRCNN уже используется только RoIAlign, помогающий сохранить оригинальные пространственные координаты, которые смещаются в случае использования RoI. Это необходимо затем, чтобы выходные данные RoIAlign можно было совместить с данными, полученными на первом этапе, и с помощью модуля Mask Head (который в свою очередь также реализован на базе сверточных слоев) сгенерировать маску для каждого ответа RoIAlign. Такие маски представляют из себя двумерную матрицу, которая для каждого пикселя, входящего в границы региона объекта, определяет, принадлежит ли этот пиксель искомому объекту или нет.

Подобный подход, с одной стороны, позволяет более точно определить границы искомого объекта. В идеальных случаях MRCNN может точно вычислить все пиксели изображения, которые отображают искомый объект. С другой стороны, в процессе обучения данные, получаемые в модуле Mask Head, используются для дополнительного обучения магистральной СНС, что позволяет повысить и ее точность работы.

При этом, поскольку для попиксельного поиска используются результаты расчета магистральной СНС, модуль Mask Head во время работы производит семантическую сегментацию только текущего искомого объекта. То есть даже если рядом расположены два объекта одного класса, Mask Head на каждой отдельной итерации своей работы будет распознавать пиксели, принадлежащие только одному из этих объектов. Такой подход позволяет MRCNN корректно решать задачу распознавания, когда искомые объекты расположены близко или даже перекрывают друг друга.

Обучение сети для распознавания северных оленей на аэрофотоснимках

Основным массивом обучения для MRCNN является массив изображений MS COCO dataset (Microsoft Common Objects in Context) [21]. На сегодня это самый крупномасштабный набор данных, используемый для обучения моделей машинного обучения решению задач обнаружения и сегментации [22, 23], состоящий из 328 тысяч изображений. Все изображения уже размечены и сформированы в обучающие выборки. Поэтому использование данного массива для базового обучения MRCNN позволяет задать для нее все основные концепции различных классов объектов, в том числе и животных. Однако изображения оленей не входят в MS COCO dataset, и MRCNN по умолчанию не способна отличить их от ряда других животных (овец, газелей, коров, лошадей).

Результаты распознавания оленей на аэрофотоснимке с помощью базовой MRCNN, обученной на MS COCO dataset, представлены на рис. 3. На снимке правильно распознанные изображения оленей помечены зеленым цветом, их четыре. Красным цветом помечены артефакты, ошибочно принятые за оленей, этих пометок 32.

Таким образом, при обучении сети необходимо учитывать особенности изображений северных оленей на аэрофотоснимках и их отличие от изображений камней, кочек и других объектов на снимках, которые могут быть приняты за оленей. Специфика аэрофотоснимков состоит в том, что фотографирование стад производится с разного расстояния, на разных ландшафтах, при различных условиях освещенности, животные на снимках имеют различную окраску и могут находиться под разными углами к камере, могут перекрывать друг друга. Эти особенности аэрофотоснимков создают дополнительные сложности при решении задачи распознавания оленей.

Задача распознавания северных оленей по аэрофотоснимкам является, таким образом, не тривиальной. Поэтому для ее решения была предложена двухэтапная процедура обучения MRCNN. Первоначальное обучение на массиве MS COCO dataset, который включает изображения других животных, необходимо для того, чтобы сеть научилась распознавать животных как класс объектов. На втором этапе MRCNN дообучается на массиве аэрофотоснимков стад северных оленей. Данный подход известен как transfer learning [24] и сегодня применяется при решении многих задач, для которых невозможно, по тем или иным причинам, собрать массив обучающих данных для конкретной рассматриваемой ситуации. Поэтому сначала модель обучается на большом массиве похожих ситуаций, а после этого уже дообучается на целевых примерах.



■ **Рис. 3.** Результат распознавания оленей с помощью базовой MRCNN, обученной на MS COCO dataset

■ **Fig. 3.** Reindeer recognition result using basic MRCNN trained on MS COCO dataset



■ **Рис. 4.** Результат распознавания оленей с помощью дообученной MRCNN
 ■ **Fig. 4.** Reindeer recognition result using the re-trained MRCNN

Был подготовлен входной массив целевых примеров, содержащий обучающую выборку из 100 аэрофотоснимков стад, на которых все животные помечены, и тестовую выборку из 30 исходных снимков стад, не включенных в обучающую выборку.

При обучении модели были установлены следующие параметры MRCNN: 20 эпох обучения; 60 шагов обучения в эпохе; скорость обучения 0,0058 (безразмерная); порог пропуска обнаружения 0,7 (безразмерный). Поскольку СНС в общем случае являются черными ящиками, нельзя в явном виде описать влияние конкретных значений параметров на результаты работы обученной сети. Данные значения были установлены исходя из практики работы с СНС и с учетом эмпирических наблюдений в процессе обучения.

После дообучения сети эксперимент по распознаванию оленей на фотоснимке, изображенном на рис. 3, был повторен. Результат эксперимента представлен на рис. 4. Модель корректно распознала на снимке 70 оленей из 93 (точность 75 %) и при этом не совершила ни одной ошибки второго рода, т. е. ни разу не приняла за оленя объект, который оленем не является. В среднем же на всем тестовом массиве данных обучения модель корректно распознала 82 % оленей.

Веб-интерфейс системы и результаты ее валидации на независимом массиве данных

На данный момент для разработанной сети создан программный комплекс с веб-интерфейсом (<https://regionview.ru/ai/>), а сама программа введена в ограниченную эксплуатацию.

Поскольку система была создана на основе сервис-ориентированной архитектуры [25], данный программный комплекс можно использовать как стационарно, так и удаленно, посредством сети Интернет.

Для работы с системой пользователь должен загрузить в свой компьютер снимки стад в форматах JPEG (.jpg) или GIF (.gif).

Интерфейс системы содержит набор оконных форм, обеспечивающих:

- загрузку аэрофотоснимков с компьютера пользователя для их обработки;
- запуск программы распознавания и подсчета оленей на снимках;
- представление результатов работы программы с отображением на экране снимка, на котором помечены распознанные системой изображения оленей, и общего числа подсчитанных животных;
- скачивание результатов на компьютер пользователя.

После просмотра помеченного снимка пользователь, если его не удовлетворит точность программного комплекса, может продолжить дальнейшую обработку снимка вручную с использованием программы ручной обработки снимков (см. разд. «Распознавание животных на аэрофотоснимках по признакам») или в любом графическом редакторе, поддерживающем расширение файла с изображением .jpg.

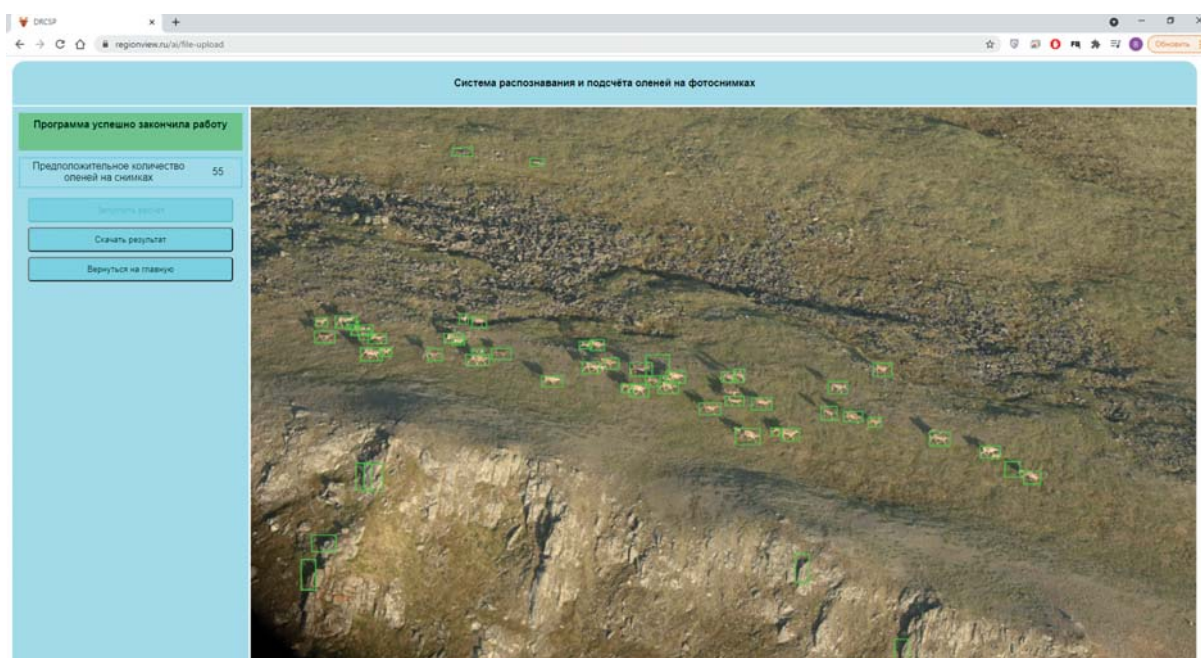
Графический интерфейс системы распознавания и подсчета оленей показан на рис. 5.

Проверка работы программы была проведена на независимом массиве из 10 аэрофотоснимков.

На фотоснимке небольшого стада оленей (рис. 6) при распознавании было совершено четыре ошибки: два оленя были распознаны дважды и два теленка не были распознаны. Всего из 41 оленя корректно было распознано 35 особей. Однако ошибка подсчета количества животных в стаде составила всего 10 %, что является достаточно высоким результатом для автоматических систем распознавания животных в естественных условиях. Причинами этого в данном случае являются высокое качество снимка и высокое разрешение самих оленей, а также сильная контрастность фона с объектами распознавания.

На снимке (рис. 7) видно, что MRCNN хорошо работает с большими стадами. При этом заметно, что сеть с одинаковой эффективностью распознает оленей, находящихся на разном удалении от камеры. Таким образом, сеть не привязана к конкретному разрешению объектов и способна работать с искаженной перспективой. Ошибка распознавания составила около 3 %.

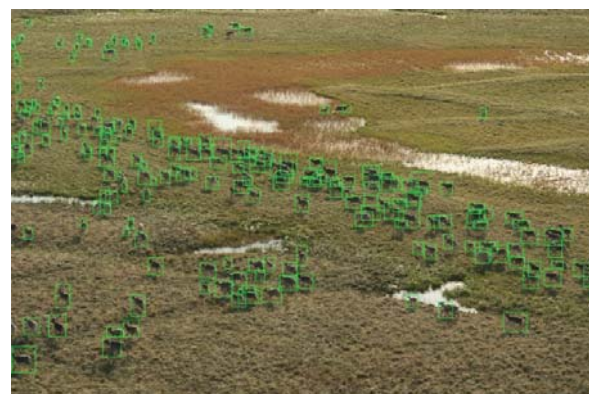
Стадо оленей на фоне типичного неоднородного тундрового пейзажа показано на рис. 8. По разметке изображений оленей видно, что MRCNN может работать со снимками, которые зашумлены фоновыми объектами — лужами, озерами, бугорками, полигонами и т. п. Объекты фона ни разу не были приняты системой распознавания за изображения оленей. Также заметно, что сеть



■ **Рис. 5.** Окно интерфейса системы, режим «Программа успешно закончила работу»
 ■ **Fig. 5.** System interface window, “Program successfully completed work” mode



■ **Рис. 6.** Распознавание оленей на однородном фоне
 ■ **Fig. 6.** Deer recognition in a homogeneous field



■ **Рис. 8.** Крупное стадо на неоднородном фоне
 ■ **Fig. 8.** Large herd against a heterogeneous background



■ **Рис. 7.** Распознавание и подсчет оленей в крупном стаде на однородном фоне
 ■ **Fig. 7.** Recognizing and counting deer in a large herd on a homogeneous background

хорошо работает со стадами, в которых олени собираются в очень плотные группы. Ошибка распознавания составила около 17 %.

Заключение

Выполненные исследования показали, что разработка автоматической системы распознавания и подсчета северных оленей в признаковом пространстве требует больших затрат времени специалистов для анализа снимков, отбора признаков и критериев разделения изображений животных от фона, проведения разного рода тестов

вых компьютерных экспериментов. СНС учится выполнять задачу классификации непосредственно по изображениям без предварительного ручного отбора признаков. Наличие готовых архитектур СНС, обученных на распознавание на изображениях различного рода объектов, дает возможность путем дообучения эффективно создавать приложения, ориентированные на распознавание конкретных видов животных. С использованием данного подхода авторами был создан программный комплекс для распознавания и подсчета северных оленей по аэрофотоснимкам, для которого можно выделить следующие перспективные направления развития.

1. Уменьшение ошибок распознавания. Необходимо увеличить объем обучающей выборки, особенно для сильно «зашумленных» и разномасштабных изображений животных. Однако связь между величиной ошибки и объемом выборки носит экспоненциально затухающий характер, а чрезмерное увеличение выборки может привести к переобучению модели. Поэтому заранее нельзя оценить количество снимков, которые потребуются добавить в обучающую выборку. Кроме того, при учете животных в скоплениях всегда может возникнуть ситуация, с которой автоматическая система распознавания справиться не сможет. В этом случае система распознавания может работать в интерактивном режиме совместно с программой обработки изображений, представленной в разд. «Распознавание животных на аэрофотоснимках по признакам». Эта программа позволяет после получения результатов автоматического распознавания корректировать их в ручном режиме, удаляя ошибочные метки и (или) ставя их на нераспознанные программой объекты.

2. Перенос программного комплекса на высокопроизводительные серверы. Специфика вычислений при использовании MRCNN накладывает ряд требований на аппаратную часть.

А коммерческое использование подразумевает параллельную работу с программным комплексом нескольких пользователей, что может существенно нагрузить систему. Несмотря на заложенную в архитектуру программного комплекса параллельность вычислений, следует провести дополнительные исследования и стресс-тесты, чтобы определить конкретные требования к аппаратной части.

3. Анализ видеоизображений. Сегодня архитектура MRCNN используется при обработке видеоизображений в ряде проектов [26]. Имеются программные библиотеки, позволяющие интегрировать данную архитектуру СНС в программные комплексы, работающие с видеопотоком. Это позволяет модифицировать представленную разработку для распознавания и подсчета оленей в видеопотоке. Подобный подход поможет ускорить процесс получения результатов учета, поскольку исключает процедуру монтажа снимков, необходимую при дискретной фотосъемке стад оленей.

4. Подсчет других видов животных и птиц. Основным массивом для начального обучения MRCNN является массив изображений MS COCO dataset. Изображения северных оленей служат для дообучения СНС. Подобным же образом, без изменения архитектуры СНС и организации интерфейса, система может быть дообучена для распознавания и подсчета стад сайгаков, стай гусей в период линьки или некоторых видов полярных чаек.

Финансовая поддержка

Работа выполнена при поддержке World-Wide Fund for Nature — Contract № BBF003107, РФФИ — грант № 19-37-90112 и бюджетной темы № 0073-2019-0004.

Литература

1. Зырянов В. А., Павлов Б. М., Якушкин Г. Д. *Экологические основы учета численности промысловых животных в тундровой зоне Таймыра*. Проблемы охотничьего хозяйства Красноярского края: материалы совещ. Красноярск, 1971. С. 70–72.
2. Kolpashikov L., Mikhailov V., Russell D. E. The role of harvest, predators, and socio-political environment in the dynamics of the Taimyr wild reindeer herd with some lessons for North America. *Ecology and Society*, 2015, vol. 20, iss. 1, article 9. <https://www.ecologyandsociety.org> (дата обращения: 13.10.20).
3. Челинцев Н. Г. *Математические основы учета животных*. М., ГУ Центрохотконтроль, 2000. 431 с.
4. Соловьев Н. В., Сергеев А. М. Распознавание на изображении множества однотипных объектов с высокой изменчивостью. *Научная сессия ГУАП: сб. докл.: в 3 ч. Ч. II: Технические науки*. СПб., ГУАП, 2019. С. 427–430.
5. Simon J. D. *Prince computer vision: Models, learning, and inference*. Cambridge, Cambridge University Press, 2012. 598 p.
6. Ayyadevara K., Reddy Y. *Modern computer vision with PyTorch: Explore deep learning concepts and implement over 50 real-world image applications*. Birmingham, Packt Publishing, 2020. 824 p.
7. Norouzzadeh M. S., Nguyen A., Kosmala M., Swanson A., Palmer M., Packer C., Clune J. Automated animal identification using deep learning techniques.

- Proc. of the Nat. Acad. of Sciences of the USA*, 2018, June, vol. 115, iss. 25, pp. 5716–5725. doi.org/10.1073/pnas.17193671155
8. Lecun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, vol. 1, iss. 4, pp. 541–551.
 9. Duporse J., Isupova O., Reece S., Macdonald D., Way T. Using very- high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *bioRxiv — the preprint server for biology*. doi: <https://doi.org/10.1002/rse2.195>
 10. *Open Data Science. Kaggle: как наши сеточки считали морских львов на Алеутских островах*. <https://habr.com/ru/company/ods/blog/337548/> (дата обращения: 20.11.20).
 11. *Seek by iNaturalist*. https://www.inaturalist.org/pages/seek_app (дата обращения: 21.11.20).
 12. Tou J., Gonzales R. *Pattern Recognition Principles*. Addison-Wesley Publishing Co, 1974. 378 p.
 13. Gonzales R., Woods R. *Digital Image Processing*. Prentis Hall, 2002. 822 p.
 14. *Программа для подсчета однотипных объектов на сжатом изображении*. Свидетельство о гос. рег. программы для ЭВМ № 2019666902 Российская Федерация/ А. М. Сергеев, Н. В. Соловьев, Г. К. Якушев. № 2019665795, заявл. 04.12.19; опубл. 17.12.19; Бюл. № 12. 1 с.
 15. Михайлов В. В., Харин Я. В. К вопросу о построении системы распознавания и подсчета животных по аэрофотоснимкам. Ч. 1. Анализ методов распознавания. *Информационно-управляющие системы*, 2011, № 2, с. 22–28.
 16. Пятаев А. С. Сегментация дерева на изображении методом неполной пороговой обработки с глобальным порогом. *Решетневские чтения*, 2018, т. 2, с. 291–293.
 17. LeCun Y., Haffner P., Bottou L., Bengio Y. *Object Recognition with Gradient-Based Learning*. Shape, Con-
tour and Grouping in Computer Vision. Springer, 1999. Pp. 319–345.
 18. He K., Gkioxari G., Dollar P., Girshick R. *Mask R-CNN*. Computer Vision and Patter Recognition. Cornell University, 2017. <https://arxiv.org/abs/1703/1703.06870>.
 19. Ganesh P., Volle K., Burks T. F., Mehta S. S. Deep orange: Mask R-CNN-based orange detection and segmentation. *IFAC-PapersOnLine*, 2019, vol. 52, iss. 30, pp. 70–75. <https://doi.org/10.1016/j.ifacol.2019.12.499>
 20. Zhao G., Hu J., Xiao W., Zou J. A mask R-CNN based method for inspecting cable brackets in aircraft. *Chinese Journal of Aeronautics*, 2020. <https://doi.org/10.1016/j.cja.2020.09.024>
 21. Lin T. Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick C. L. Microsoft COCO: Common objects in context. *Computer Vision — ECCV 2014*, 2014, part 5, pp. 740–755.
 22. Patterson G., Hays J. COCO attributes: Attributes for people, animals, and objects. *Computer Vision — ECCV 2016*, 2016, part 6, pp. 85–100.
 23. Srivastava S., Divekar A. V., Anilkumar C., Naik I., Kulkarni V., Pattabiraman V. Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, 2021, no. 8, article 66. doi:10.1186/s40537-021-00434-w
 24. Weiss K., Khoshgoftaar T. M., Wang D. D. A survey of transfer learning. *Journal of Big Data*, 2016, vol. 3, article 9. doi:10.1186/s40537-016-0043-6
 25. Niknejad N., Ismail W., Ghani I., Nazari B., Bahari M., Ab Razak Bin Che Hussin. Understanding Service-Oriented Architecture (SOA): A systematic literature review and directions for further investigation. *Information Systems*, 2020, vol. 91, article 101491.
 26. Yang L., Fan Y., Xu N. Video instance segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 5188–5197.

UDC 004.932.2

doi:10.31799/1684-8853-2021-5-20-32

Methodological approaches and algorithms for recognizing and counting animals in aerial photographsV. V. Mikhailov^a, Dr. Sc., Tech., Professor, Leading Researcher, orcid.org/0000-0001-6973-3526, mwwcari@gmail.comL. A. Kolpaschikov^b, Dr. Sc., Biol., Head of a Research Division, orcid.org/0000-0003-4871-3360B. A. Sobolevskii^a, Junior Researcher, Post-Graduate Student, orcid.org/0000-0001-7685-4991N. V. Soloviev^c, PhD, Tech., Associate Professor, orcid.org/0000-0001-8851-3713G. K. Yakushev^c, Bachelor, orcid.org/0000-0003-4689-3425^aSt. Petersburg Federal Research Center of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation^bDirectorate of Taimyr Nature Reserves, Talnakhskaia St., 22, entrance 2, 663300, Norilsk, Russian Federation^cSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaja St., 190000, Saint-Petersburg, Russian Federation

Introduction: The complexity of recognition and counting of objects in a photographic image is directly related to variability of related factors: physical difference of objects from the same class, presence of images similar to objects to be recognized, non-uniform background, change of shooting conditions and position of the objects when the photo was taken. Most challenging are the problems of

identifying people in crowds, animals in natural environment, cars from surveillance cameras, objects of construction and infrastructure on aerial photo images, etc. These problems have their own specific factor space, but the methodological approaches to their solution are similar. **Purpose:** The development of methodologies and software implementations solving the problem of recognition and counting of objects with high variability, on the example of reindeer recognition in the natural environment. **Methods:** Two approaches are investigated: feature-based recognition based on binary pixel classification and reference-based recognition using convolutional neural networks. **Results:** Methodologies and programs have been developed for pixel-by-pixel recognition with subsequent binarization, image clustering and cluster counting and image recognition using the convolutional neural network of Mask R-CNN architecture. The network is first trained to recognize animals as a class from the array of MS COCO dataset images and then trained on the array of aerial photographs of reindeer herds. Analysis of the results shows that feature-based methods with pixel-by-pixel recognition give good results on relatively simple images (recognition error 10–15%). The presence of artifacts on the image that are close to the characteristics of the reindeer images leads to a significant increase in the error. The convolutional neural network showed higher accuracy, which on the test sample was 82%, with no false positives. **Practical relevance:** A software prototype has been created for the recognition system based on convolutional neural networks with a web interface, and the program itself has been put into limited operation.

Keywords — recognition, convolutional neural networks, aerial images, reindeer.

For citation: Mikhailov V. V., Kolpaschikov L. A., Sobolevskii B. A., Soloviev N. V., Yakushev G. K. Methodological approaches and algorithms for recognizing and counting animals in aerial photographs. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 5, pp. 20–32 (In Russian). doi:10.31799/1684-8853-2021-5-20-32

References

- Zirjanov V. A., Pavlov B. M., Jakushkin G. D. *Ekologitceskie osnovi utcheta chislennosti promislovih zivotnih v tundrovoi zone Taimyra*. In: *Problemi ohotnichogo hoziaistva Krasnoyarskogo kpaia* [Ecological basics of counting the number of hunting animals in the tundra zone of Taimyr. In: Meeting materials “Problems of hunting in Krasnoyarsk region”]. Krasnoyarsk, 1971. Pp. 70–72 (In Russian).
- Kolpashikov L., Mikhailov V., and Russell D. E. The role of harvest, predators, and socio-political environment in the dynamics of the Taimyr wild reindeer herd with some lessons for North America. *Ecology and Society*, 2015, vol. 20, iss. 1, article 9. Available at: <https://www.ecologyandsociety.org> (accessed 13 October 2020).
- Chelintsev N. G. *Matematicheskie osnovi utchea zivotnih* [Mathematical basis of animal counting]. Moscow, GU Centrohtkontrol Publ., 2000. 431 p. (In Russian).
- Soloviev N. V., Sergeev M. B. Image recognition of a set of homogeneous objects with high variability. *Nautshnaia sessija GUAP: sbornik dokladov: v 3 chastiah. Chast II: Tehniceskie Nauki* [Scientific session of SUAE: collection of reports: in 3 parts. Part II: Technical Sciences]. Saint-Petersburg, GUAP Publ., 2019, pp. 427–430 (In Russian).
- Simon J. D. *Prince computer vision: Models, learning, and inference*. Cambridge, Cambridge University Press, 2012. 598 p.
- Ayyadevara K., Reddy Y. *Modern computer vision with PyTorch: Explore deep learning concepts and implement over 50 real-world image applications*. Birmingham, Packt Publishing, 2020. 824 p.
- Norouzzadeh M. S., Nguyen A., Kosmala M., Swanson A., Palmer M., Packer C., Clune J. Automated animal identification using deep learning techniques. *Proc. of the Nat. Acad. of Sciences of the USA*, 2018, June, vol. 115, iss. 25, pp. 5716–5725. doi.org/10.1073/pnas.17193671155
- Lecun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, vol. 1, iss. 4, pp. 541–551.
- Duporse J., Isupova O., Reece S., Macdonald D., Way T. Using very high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *bioRxiv — the preprint server for biology*. doi: <https://doi.org/10.1002/rse2.195>
- Open Data Science. Kaggle: kak nashi setochki schitali morskih l'vov na Aleutskih ostrovah* [Open Data Science. Kaggle: how our nets counted sea lions on Aleutian Islands]. Available at: <https://habr.com/ru/company/ods/blog/337548/> (accessed 20 November 2020) (In Russian).
- Seek by iNaturalist*. Available at: https://www.inaturalist.org/pages/seek_app (accessed 21 November 2020).
- Tou J., Gonzales R. *Pattern recognition principles*. Addison-Wesley Publishing Co, 1974. 378 p.
- Gonzales R., Woods R. *Digital image processing*. Prentice Hall, 2002. 822 p.
- Sergeev M. B., et al. *Programma dla podscheta odnotypnih ob'ektov na shtom izobrajenii* [A programme for counting of similar objects in a compressed image]. State Registration Certificate for a Computer Programme, no. 2019666902, 2019.
- Mikhailov V. V., Harin Y. V. On the developing of an animal recognition and counting system for aerial photographs. 1. Analysis of recognition methods. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2011, no. 2, pp. 22–28 (In Russian).
- Plataea A. S. Tree segmentation on an image using an incomplete thresholding method with a global threshold. *Reshetnevskie chteniya*, 2018, vol. 2, pp. 291–293 (In Russian).
- LeCun Y., Haffner P., Bottou L., Bengio Y. *Object Recognition with Gradient-Based Learning*. In: *Shape, Contour and Grouping in Computer Vision*. Springer, 1999. Pp. 319–345.
- He K., Gkioxari G., Dollar P., Girshick R. *Mask R-CNN*. In: *Computer Vision and Patter Recognition*. Cornell University, 2017. <https://arxiv.org/abs/1703.06870>.
- Ganesh P., Volle K., Burks T. F., Mehta S. S. Deep orange: Mask R-CNN-based orange detection and segmentation. *IF-AC-PapersOnLine*, 2019, vol. 52, iss. 30, pp. 70–75. <https://doi.org/10.1016/j.ifacol.2019.12.499>
- Zhao G., Hu J., Xiao W., Zou J. A mask R-CNN based method for inspecting cable brackets in aircraft. *Chinese Journal of Aeronautics*, 2020. <https://doi.org/10.1016/j.cja.2020.09.024>
- Lin T. Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick C. L. Microsoft COCO: Common objects in context. *Computer Vision — ECCV 2014*, 2014, part 5, pp. 740–755.
- Patterson G., Hays J. COCO attributes: Attributes for people, animals, and objects. *Computer Vision — ECCV 2016*, 2016, part 6, pp. 85–100.
- Srivastava S., Divekar A. V., Anilkumar C., Naik I., Kulkarni V., Pattabiraman V. Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, 2021, no. 8, article 66. doi:10.1186/s40537-021-00434-w
- Weiss K., Khoshgoftaar T. M., Wang D. D. A survey of transfer learning. *Journal of Big Data*, 2016, vol. 3, article 9. doi:10.1186/s40537-016-0043-6
- Niknejad N., Ismail W., Ghani I., Nazari B., Bahari M., Ab Razak Bin Che Hussin. Understanding Service-Oriented Architecture (SOA): A systematic literature review and directions for further investigation. *Information Systems*, 2020, vol. 91, article 101491.
- Yang L., Fan Y., Xu N. Video instance segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 5188–5197.