

СИСТЕМА АНАЛИЗА ДАННЫХ С ФОТОЛОВУШЕК ДЛЯ ОПЕРАТИВНОГО ДИСТАНЦИОННОГО МОНИТОРИНГА ПРИРОДНЫХ ТЕРРИТОРИЙ

В.А. Ефремов

А.В. Леус

Д.А. Гаврилов

Д.И. Мангазеев

И.В. Холодняк

А.С. Радыш

В.А. Зуев

Н.А. Водичев

М.М. Паршиков

efremov.va@phystech.edu

leus.av@mipt.ru

gavrilov.da@mipt.ru

mangazeev.di@phystech.edu

kholodnyak.iv@phystech.edu

radysh.as@phystech.edu

zuev.va@phystech.edu

vodichev.na@phystech.edu

parchikov.mm@phystech.edu

МФТИ, г. Долгопрудный, Московская обл., Российская Федерация

Аннотация

Приведена система анализа данных, полученных с помощью фотоловушек, для оперативного дистанционного мониторинга природных территорий на основе двухстадийной нейросетевой обработки изображений, содержащая серверную и пользовательскую части. Серверная часть предназначена для обработки большого объема данных, полученных из разных заповедников в целях обучения нейросетевых алгоритмов. Пользовательская часть необходима для установки на локальный вычислитель в заповеднике. Разработанная система позволяет значительно сократить время обработки данных и упрощает экологический анализ, возможность дообучения классификатора под видовое разнообразие любого заповедника без переобучения детектора, повышает качество распознавания видов животных в пределах одной особо охраняемой природной территории, что делает систему более гибкой и масштабируемой. Для корректировки количественных и качественных предсказаний алгоритмов программное обеспечение дополнено функциональностью, которая позволяет автоматически создавать так называемые регистрации. С помощью регистраций ведется подсчет числа объектов на каждом фотоизображении с учетом

Ключевые слова

Изображения с фотоловушек, агрегативная кластеризация, глубокие сверточные нейронные сети, детекция, классификация, двухстадийный подход, регистрации

контекстной информации из их последовательности. Регистрации позволяют корректировать предсказания нейронной сети не только по числу животных на фотоизображении, но и по предсказанным классам. Проведено сравнение скорости работы системы на различных аппаратных платформах. Показано, что использование современных графических вычислителей позволяет осуществлять обработку фото- и видеоизображений со скоростью, значительно превосходящей возможности человека

Поступила 21.02.2023

Принята 27.03.2023

© Автор(ы), 2023

Введение. Для автоматической съемки животных в дикой природе широко используется такой инструмент, как фотоловушки, позволяющие проводить мониторинг больших участков природных территорий [1, 2]. Общим недостатком использования фотоловушек является необходимость накопления и передачи избыточной информации [3, 4]. При активном использовании фотоловушек объем данных, накапливаемый за год, может составлять более 200 ГБ только в пределах одной особо охраняемой природной территории. Передача всего массива сырых необработанных данных, содержащего большое число избыточных фото- и видеоизображений, записей с избыточным качеством или разрешением, а также пустых файлов, которые получены в результате ложного срабатывания фотоловушки, является низкоэффективным решением.

Технологии компьютерного зрения набирают широкую популярность и активно используются в задачах детектирования и классификации [5, 6], в том числе для мониторинга животных в заповедниках [7–10]. Архитектура нейронной сети, позволившая превзойти человеческие возможности в качестве распознавания фото- и видеоизображений, описана в [11]. Глубокие сверточные нейронные сети открывают путь к автоматизации процесса обработки данных, полученных с помощью фотоловушек [12–15]. Возможности по классификации и подсчету особей с помощью нейронных сетей *Faster R-CNN* и *YOLOv2* на датасетах *RCT* и *GSSS* приведены в [16]. Лучшие результаты показал алгоритм *Faster R-CNN* — точности детектирования и подсчета составили $93 \pm 3,2$ % и $76,7 \pm 8,31$ %. Проблема обобщения моделей к новым локациям, на которых ставятся новые фотоловушки, на датасете *CCT* (*Caltech Camtra Traps*) описана в [17]. Показано, что метрики качества сильно ухудшаются на фотоизображениях с фотоловушек, на которых не происходило обучение модели. Техника активного

обучения для разметки большого набора данных с применением обучения вложений рассмотрена в [18]. Кроме того, использован двухстадийный подход в распознавании животных, где на первой стадии выполнялась детекция объектов, чтобы отделить животное от фона и подсчитать число особей на фотоизображении, а затем классификация, чтобы определить вид особи. Приложение глубоких нейронных сетей для экологического анализа, а именно для подсчета видового богатства и выделения паттернов активности и занятости, приведено в [19].

В настоящей работе разработана система анализа данных с фотоловушек для оперативного дистанционного мониторинга природных территорий. *Цель настоящей работы* — разработка технологии анализа данных, полученных с помощью фотоловушек, на основе двухстадийной глубокой нейросетевой обработки. Для достижения поставленной цели решены следующие задачи:

- сбор и разметка полученных данных, подготовка обучающих выборок;
- разработка алгоритма поиска и удаления изображений-дубликатов, адаптация разработанного алгоритма для разделения выборки на обучающую, валидационную и тестовую;
- выбор наилучшего детектора объектов с точки зрения качества и скорости работы на наборах данных, собранных в заповедниках России;
- выбор оптимального классификатора, способного с высокой точностью распознавать обнаруженные детектором объекты;
- программная реализация технологии анализа данных на основе алгоритмов, показавших лучшие результаты;
- оценка скорости разработанной системы на основе двухстадийного нейросетевого подхода на различных аппаратных платформах;
- повышение количественных и качественных результатов предсказаний двухстадийного нейросетевого подхода к фото- и видеоизображениям с фотоловушек с помощью регистраций.

Методы исследования. Сбор данных. Для проведения исследований собраны данные с фотоловушек из 51 заповедника, расположенных на территории России. Изначальная выборка имела сильный дисбаланс классов, не структурирована и сильно загрязнена, фотоизображения часто дублируются. Общий объем данных составил 3 ТБ, из которых 1 млн фотоизображений и 65 тыс. видеоизображений. Фото- и видеоизображения собирались в различных климатических зонах при разных заднем фоне, погодных условиях, освещении, времени суток и т. д. Указанные фото- и видеоизоб-

ражения размечены с присвоением класса. Время разметки 231 тыс. изображений для задачи детекции и 416 тыс. изображений для задачи классификации составило примерно один месяц. Каждое видеоизображение нарезалось на кадры, что позволило пополнить обучающую выборку. Помимо основного датасета использовался публично открытый датасет, на котором проверялась работоспособность детектора и классификатора, состоящий из 7600 изображений для детекции и 7000 изображений для классификации. В качестве основных классов выступают тигры и леопарды. Из-за недостатка некоторых видов животных в выборке (белки, рыси, лисы, собаки, благородного оленя) недостающие классы дополнены с помощью открытого датасета *NACIT* [15], содержащего 3,7 млн изображений. Выбор в пользу датасета *NACIT* сделан из-за того, что он получен в условиях окружающей среды, наиболее схожих с условиями в заповедниках России.

Удаление фото и видеоизображений-дубликатов. С помощью специальной программы, написанной в среде MATLAB, выполнялся сравнительный анализ изображений на схожесть в рамках одного заповедника. Такой подход позволяет удалять из выборки одинаковые или практически неотличимые фото- и видеоизображения. Классические способы, рассчитывающие корреляции между изображениями, вычислительно недостаточно эффективны, поскольку приходится сравнивать каждое изображение со всеми. Разработанный подход позволяет в несколько раз ускорить процесс поиска фото- и видеоизображений-дубликатов. Схема подхода имеет следующий вид.

1. Для одного изображения данного заповедника выполняем следующие действия:

- приводим к оттенкам серого и применяем дискретное косинусное преобразование;
- из полученной матрицы выбираем первые восемь элементов по вертикали и горизонтали, поскольку в них сосредоточены самые низкие частоты; полученную матрицу 8×8 приводим к 64-мерному вектору (*vec*);
- вычисляем среднее значение по всему вектору (vec_{mean});
- каждое значение в векторе сравнивается с vec_{mean} ; если значение элемента вектора больше vec_{mean} , то присваиваем ему значение единицы, в противном случае — нуля. Таким образом, получаем *p-hash* (перцептивное хеширование);
- исходное изображение приводим к размеру 4×4 ;

– вычисляем разность яркости между левой и правой (*left / right*) и верхней и нижней (*top / bottom*) частями изображения 4×4 ;

– в общую таблицу добавляем запись со значениями *mean_br_{current}*, *vec*, *left / right*, *top / bottom*, где *mean_br_{current}* — среднее значение яркости текущего изображения.

2. Для текущей записи в таблице реализуем следующее действие:

– проверить выполнение условий

$$\begin{aligned} & |mean_br_{all} - mean_br_{current}| < \\ & < \langle br_{threshold} |left / right_{all} - left / right_{current}| \times \\ & \times \langle left / right_{threshold} |top / bottom_{all} - top / bottom_{current}| < \\ & < top / bottom_{threshold}, \end{aligned}$$

где *mean_br_{all}* — вектор средних значений яркости для всех фото- и видеоизображений; *left / right_{all}* и *top / bottom_{all}* — векторы разности яркости между левой и правой, верхней и нижней частями изображения для всех фото- и видеоизображений; *br_{threshold}*, *left / right_{threshold}*, *top / bottom_{threshold}* — пороговые значения яркости (устанавливаются экспериментально).

Для решения задачи детекции алгоритм применяем к данным в целях удаления изображений-дубликатов с порогом корреляции 99 %. Выбор такого высокого порога корреляции обусловлен тем, что на изображениях с далеко расположенными животными, сделанных с интервалом в 1 с, могут наблюдаться существенные изменения позы. При этом задний фон практически не изменяется. При установлении порога корреляции менее 99 % такие изображения будут считаться одинаковыми и удаляться, что нежелательно.

Формирование выборки. Для разделения выборок на обучающую, валидационную и тестовую, подход по удалению дубликатов объединен с методом агломеративной кластеризации, с помощью которого 64-мерные хеш-векторы изображений объединялись в кластеры. Расстояние между хеш-векторами рассчитывали с помощью евклидовой метрики. Данный подход позволяет разделять максимально непохожие изображения по разным кластерам. Затем с помощью алгоритма поиска дубликатов сравнивают, насколько сильно отличаются между собой выборки для обучения, валидации и тестирования, полученные при случайном разделении изображений и с помощью кластерного подхода. Датасеты для задач детекции и классификации разделяют в соотношениях 85, 9, 6 % и 77, 15, 8 %,

где первое, второе и третье число — относительные объемы обучающей, валидационной и тестовой выборок соответственно.

Метрики для задач детекции и классификации. Для оценки качества работы методов детекции использованы метрика *IOU (Intersection over Union)*, позволяющая рассчитать степень перекрытия между истинной и предсказанной ограничивающими рамками, и метрика *Average Precision (AP)*, которую можно определить как площадь под кривой *Precision–Recall*.

Для оценки качества работы алгоритмов классификации использовалась доля правильных ответов (*Accuracy*), *Precision*, *Recall* и *F1*, где

$$F1 = \frac{2Precision \cdot Recall}{Precision + Recall}.$$

Метод решения задачи детекции. Обучение детекторов для отделения фона изображения от животного является ключевой задачей настоящего исследования. Датасет для детекции размечен в полуавтоматическом режиме с использованием разработанного графического приложения с обученным алгоритмом *YOLOv5* на изображениях с разрешением 640×480 , полученных с помощью фотоловушек, установленных в Центральном-Лесном государственном природном биосферном заповеднике (Тверская обл.).

Для проведения вычислительных экспериментов выбраны алгоритмы, основанные на глубоких нейронных сетях *YOLO–YOLOv5* [20], *YOLOv5* [21], *YOLOX* [22]. Все нейронные сети предобучены на датасете *COCO (Common Objects in Context)* [23]. Параметры для сетей *YOLOv5*, *YOLOv5* и *YOLOX* выбраны по рекомендациям их авторов. Нейронные сети обучались в течение 40 эпох на трех графических процессорных устройствах (ГПУ) *GeForce RTX 3080* (10 ГБ).

Метод решения задачи классификации. Для решения задачи классификации сформирован набор данных, преимущественно состоящий из обрезанных изображений из датасета для задачи детекции и изображений животных из интернета. В качестве классификаторов выбраны модифицированные нейронные сети поколений *ResNet: SeResNet* [24], *ResNeSt* [25], *ReXNet* [26]. Выбор обусловлен тем, что модель *ResNet* показала лучший результат в задаче распознавания видов животных [12, 13, 15, 27]. Данные архитектуры сравнивались с архитектурой *EfficientNetV2* [28], которая по качеству распознавания на датасете *ImageNet* лучше архитектуры *ResNeSt* и предыдущей архитектуры *EfficientNetV1* [29]. Эффективность достигается за счет определения оптимального соотношения слоев *Fused–MBConv*

и *MVConv* с помощью алгоритма *NAS* [30], применения неравномерного масштабирования сети и техники постепенного обучения, которая позволяет сокращать время обучения и повышает точность алгоритма. В настоящей работе авторы отказались от техники постепенного обучения.

Весы нейронов для всех указанных архитектур предобучены на датасете *ImageNet*. При этом у каждой архитектуры «размораживались» все слои. Обучение моделей проводилось с помощью стохастического градиентного спуска в течение 35 эпох с коэффициентом обучения 0,0003, моментом 0,9, размером батча 16, *EMA* сглаживанием весов и с техникой сглаживания меток (*Label Smoothing*) [31]. Для увеличения числа обучающих изображений использовалась аугментация данных, например, зеркальное отражение по горизонтали (*Horizontal Flip*), случайные повороты в пределах 20° (*Random Rotation*). Обучение проводилось на одной *GeForce RTX 3080* (10 ГБ).

Вычислительные эксперименты. Разделение данных и поиск дубликатов. Для оценки работоспособности предложенного подхода использована выборка, состоящая из 10 000 изображений рысей. Разделение выборки на обучающую, валидационную и тестовую осуществлено с помощью методов агломеративной кластеризации [32] и случайного разделения. Для сравнения обучающих, валидационных и тестовых наборов данных, полученных с помощью разных подходов к разделению, использован алгоритм поиска дубликатов с разными значениями порога корреляции. Общее число найденных дубликатов между выборками приведено в табл. 1.

Согласно данным табл. 1 при пороге корреляции 0,95 доля похожих изображений в случайно разделенной выборке с рысями составляет 0,2 %. При уменьшении порога корреляции до 0,8 доля похожих изображений увеличивается до 8,69 % в отличие от кластерного разделения данных, при котором дубликатов в 4 раза меньше. За счет агломеративной кластеризации данных удастся создать отличающиеся друг от друга датасеты для обучения и тестирования, что позволяет нейронной сети обучаться более качественно и быть более устойчивой к новым данным.

Эксперименты по детекции объектов. Архитектуры *YOLOv5-M6/L6/X* и *YOLOv5-W6* обучены на разрешении 1280, архитектуры *YOLOv5-X* и *YOLOv5-X* — на разрешении 640. Более сложные архитектуры такие, как *YOLOv5-X6*, *YOLOv5-E6*, *YOLOv5-D6* не рассматривались, поскольку обучение таких архитектур занимает слишком много времени на используемых ГПУ. Кроме того, большинство заповедников в России имеют ограниченные вычислительные ресурсы, поэтому инференс на более сложных архитектурах потребует больше времени для обработки фото- и видеоизображений.

Таблица 1

Результаты определения изображений-дубликатов между разными выборками

Данные, полученные с фотоловушек	Порог корреляции	Обучающие и валидационные данные			Обучающие и тестовые данные			Валидационные и тестовые данные	Число/доля изображений-дубликатов
		7	1	1	1	1	1		
<i>Случайное разделение данных</i>									
Изображения	0,95	15	2	2	2	2	2	19/0,19 %	
Изображения и дубликаты		54	22	22	3	3	3		
Изображения	0,9	117	47	47	6	6	6	170/1,7 %	
Изображения и дубликаты		123	56	56	14	14	14		
Изображения	0,85	284	129	129	26	26	26	421/4,21 %	
Изображения и дубликаты		244	102	102	36	36	36		
Изображения	0,8	624	309	309	83	83	83	869/8,69 %	
Изображения и дубликаты									
<i>Кластерное разделение данных</i>									
Изображения	0,95	0	0	0	0	0	0	0	
Изображения и дубликаты		0	0	0	0	0	0		
Изображения	0,9	3	2	2	1	1	1	12/0,12 %	
Изображения и дубликаты		6	4	4	2	2	2		
Изображения	0,85	13	13	13	4	4	4	67/0,67 %	
Изображения и дубликаты		31	30	30	8	8	8		
Изображения	0,8	35	36	36	13	13	13	207/2,07 %	
Изображения и дубликаты		93	108	108	26	26	26		

Оценки метрик качества детекции на валидационном (AP_{50}^{val} , $AP_{50:95}^{val}$) и тестовом (AP_{50}^{test} , $AP_{50:95}^{test}$) датасетах приведены в табл. 2.

Таблица 2

**Оценка метрик качества детекторов на валидационном
и тестовом датасетах**

Model	Params	Train resolution	Inference resolution	AP_{50}^{val}	$AP_{50:95}^{val}$	AP_{50}^{test}	$AP_{50:95}^{test}$	FPS	Training time, h
YOLOv5-M6	35,2	1280	1280	97	82,8	98,4	85,1	42	38
			640	96,3	80,4	98,1	83	71	
YOLOv5-L6	76,1	1280	1280	97,1	83,1	98,5	85,5	35	74
			640	96,7	81,2	98,3	83,5	58	
YOLOR-W6	79,2	1280	1280	96,7	80,1	98,1	82,6	33	107
			640	96,1	77,7	97,6	80,3	44	
YOLOv5-X	86	1280	1280	95,9	76,1	96,9	76,9	27	123
			640	96,4	78,1	97,7	79,2	64	
		640	1280	92,3	69,4	93,8	70,7	27	39
			640	97	82,7	98,6	85	64	
YOLOX-X	99,1	640	640	96	78,3	97,5	80,6	35	85

Примечание: Params — число параметров нейронной сети; Train resolution, Inference resolution — разрешения изображений на обучении и инференсе; AP_{50} , $AP_{50:95}$ — метрики Average Precision при $IOU = 0,5$ и $IOU = 0,5-0,95$; FPS — число обрабатываемых изображений на инференсе за 1 с при размере батча 1 на ГПУ GeForce RTX 3080; Training time (h) — время обучения модели (в часах).

Результаты показывают, что на валидационной выборке лучше всех себя показала архитектура YOLOv5-L6 с $AP_{50}^{val} = 97,1$ %, в то время как на тестовой выборке архитектура YOLOv5-X, обученная на разрешении 640, показала результаты выше, чем архитектура YOLOv5-L6 на 0,1 % по метрике AP_{50}^{test} . В качестве финальной выбрана архитектура YOLOv5-L6, поскольку она имеет лучшее соотношение между качеством и скоростью работы среди детекторов, обученных на разрешении 1280. Кроме того, детекторы, обученные на разрешении 1280, с большей вероятностью находят мелких особей животных в отличие от детекторов, обученных на разрешении 640.

Эксперименты по классификации объектов. Для финальной оценки качества классификации на валидационных и тестовых датасетах все изображения приводились к размеру 320×320 . В табл. 3 приведены результаты оценки эффективности рассматриваемых архитектур по метрикам *Precision*, *Recall*, *F1* на валидационном (P^{val} , R^{val} , $F1^{val}$) и тестовом (P^{test} , R^{test} , $F1^{test}$) датасетах.

Таблица 3

**Оценка метрик качества классификаторов
на валидационных и тестовых датасетах**

Model	Params	Train resolution	P^{val}	R^{val}	$F1^{val}$	P^{test}	R^{test}	$F1^{test}$	Training time, h
<i>EfficientNetV2-m</i>	53,2 М	320×320	97,99	98,3	98,11	97,98	98,48	98,18	40
<i>SeResNet-152</i>	66,8 М	256×256	98,52	97,63	98,04	98,51	98,107	98,3	44
<i>ResNeSt-101</i>	48,3 М	256×256	98,31	98,3	98,30	98,18	98,53	98,34	42
<i>ReXNet-100</i>	48 М	224×224	98,01	97,35	97,66	98,08	97,37	97,69	15
<i>ResNet-101</i>	44,6 М	256×256	98,26	96,88	97,47	98,03	96,94	97,41	33

Выбор итоговой архитектуры осуществлялся по метрике *F1*, поскольку метрика Ассигасу неустойчива к несбалансированным данным. Результаты экспериментов показали, что в задаче распознавания видов животных лучшей является архитектура *ResNeSt-101* по метрике *F1*, равная 98,3 и 98,34 % на валидационных и тестовых данных. Базовая архитектура *ResNet-101* имеет самую низкую оценку *F1*, равную 97,47 и 97,41 % на валидационных и тестовых данных.

Из табл. 3 следует, что время обучения классификатора на одном ГПУ в 3–9 раз меньше, чем время обучения детектора на трех ГПУ. Поскольку переобучение детектора для каждого заповедника требует значительных вычислительных ресурсов, но при этом не дает существенного прироста в качестве распознавания, принято решение добавить в разработанную систему функциональность, позволяющую для каждого заповедника обучать только классификатор и использовать единый универсальный детектор.

Реализация системы. Система анализа данных, полученных с помощью фотоловушек, для оперативного дистанционного мониторинга природных территорий содержит серверную и пользовательскую части. Серверная часть предназначена для обработки большого объема данных, полученных из разных заповедников в целях обучения универсальных нейросетевых алгоритмов. Пользовательская часть необходима для уста-

новки на локальный вычислитель. С помощью пользовательской части выполняется переобучение/дообучение нейросетей с учетом классов животных, необходимых заповеднику, а также последующая обработка фото- и видеоизображений, собранных в заповеднике. Серверная часть включает в себя два модуля:

1) модуль обучения универсального детектора, с помощью которого можно по размеченным данным обучить детектор, способный на фотоизображениях обнаруживать животных, людей и технику (без детальной классификации);

2) модуль обучения классификатора, с помощью которого по размеченным данным можно обучить классификатор, способный разделить найденных детектором животных на необходимое число классов.

Пользовательская часть также включает в себя два модуля:

– модуль дообучения/переобучения классификатора, который позволяет настроить универсальный классификатор под требования данного заповедника;

– модуль автоматической обработки данных, который позволяет с помощью универсального детектора (ранее обученного на мощном сервере и адаптированного под местное видовое разнообразие) находить животных в потоке или архиве фото- и видеоизображений и классифицировать этих животных.

Техническая архитектура системы анализа данных для оперативного дистанционного мониторинга природных территорий приведена на рис. 1.

Для реализации технологии анализа данных выполнено объединение нейронных сетей *YOLOv5-L6* и *ResNeSt-101*, показавших лучшие результаты, что позволило создать двухстадийную систему обработки фото- и видеоизображений, полученных с помощью фотоловушек. Для реализации функции детектирования объекта оригинальное изображение или его нормированный фрагмент видимого диапазона подается на глубокую сверточную нейросеть. На выходе получается изображение с набором координат ограничивающих рамок (*bounding box*), с помощью которых выделены все объекты интереса. Полученные координаты этих объектов добавляются в базу координат объектов. Далее каждое изображение анализируется с помощью нейросетевого классификатора, который присваивает класс каждому найденному объекту в соответствии с перечнем классов данного заповедника. Информация о классах также сохраняется в базу данных. На основе данных, содержащихся в метаданных *EXIF*, при постобработке учитывается информация о дате и времени получения этого фото- и или видеоизображения. На основе результатов анализа классов



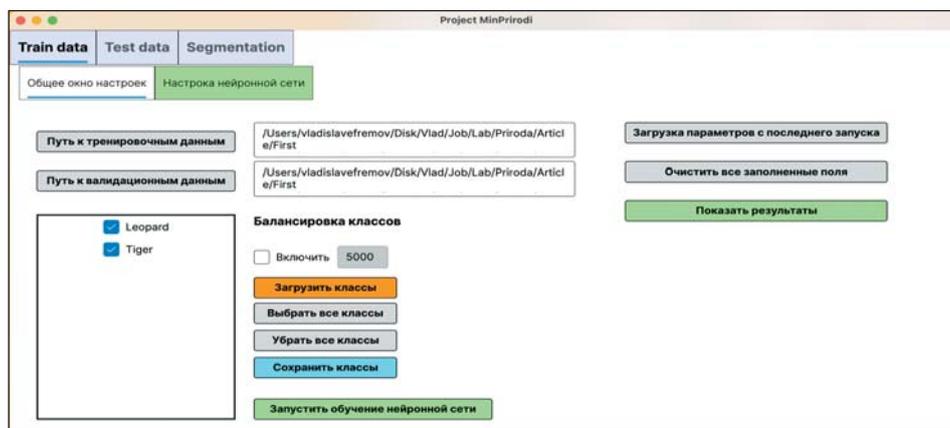
Рис. 1. Техническая архитектура системы анализа данных для оперативного дистанционного мониторинга природных территорий

в совокупности с временными и пространственными данными об изображении формируется отчет об обнаружении тех или иных объектов перед фотоловушкой.

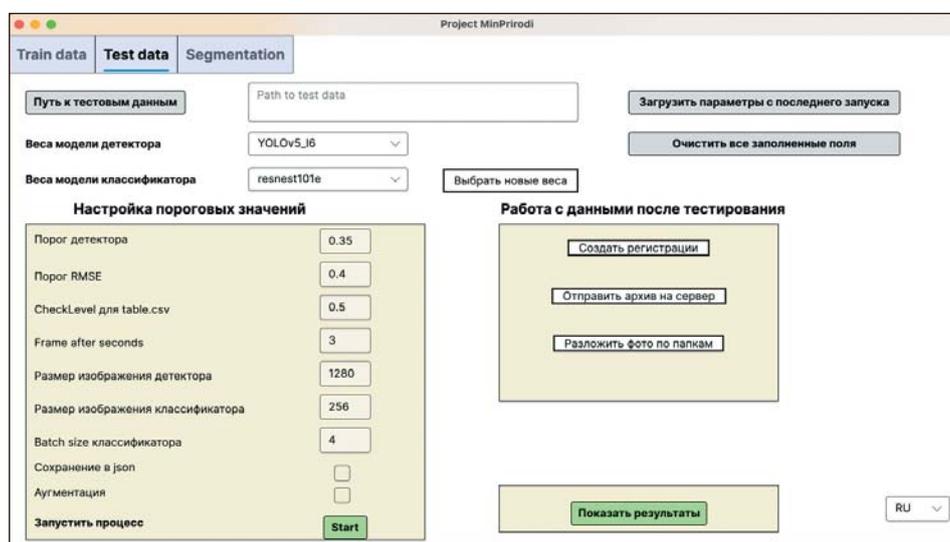
На вход нейронной сети обнаружения подается нормированный фрагмент оригинального изображения видимого и (или) дальнего ИК-диапазона. При необходимости изображение может быть сжато или фрагментировано для наиболее эффективной обработки с помощью нейросетевого детектора. Результатом обработки является рамка, обведенная вокруг найденного объекта. Полученные координаты объектов интереса добавляются в базу объектов. Из исходного несжатого изображения выделяется изображение объекта для последующей классификации.

Для реализации функции классификации на вход классифицирующей нейронной сети подаются фрагменты оригинальных изображений видимого диапазона размером 320×320 пикселей, содержащие объекты интереса. На выходе обеспечивается визуализация соответствующего типа объекта рамкой и текстом.

Программное обеспечение системы реализовано с помощью фреймворка *Avalonia UI* и может быть развернуто на операционных системах *Windows 10*, *Linux*, *MacOS*, *Raspbian* и др. Окна графического интерфейса приведены на рис. 2.



a



b

Рис. 2. Окна графического интерфейса:

- a* — для обучения классификатора под видовое разнообразие конкретного заповедника;
- b* — для обработки фото- и видеоизображений с помощью двухстадийного нейросетевого подхода

У пользователя имеется возможность указать путь до директории с фото- и видеоизображениями, которую он хочет обработать, а также выбрать веса предварительно обученного детектора и классификатора. Для более продвинутых пользователей есть возможность проводить тонкую настройку инференса, изменяя, например размер изображения, размер батча, порог детектора и др.

Оценка скорости работы системы на различных аппаратных платформах. Проведены экспериментальные исследования по обработке

1000 фото- и 100 видеоизображений на различных аппаратных платформах. Как показывает практика, заповедники имеют ограниченные вычислительные ресурсы, поэтому важно, чтобы система обрабатывала фотоизображения со скоростью выше, чем это делает работник заповедника вручную. Результаты экспериментов приведены в табл. 4.

Таблица 4

Оценка скорости работы графического приложения, установленного на компьютеры с различной конфигурацией

Центральный процессор	Операционная система	Графическая карта	Размер изображения	Время обработки	
				1000 фотоизображений	100 видеоизображений
M1	MacOS	–	640	53 мин	46 мин
Intel Core i7-6700HQ CPU 2.6GHz	Linux (Fedora)	GTX 960M (2 Гб, notebook)	1280	26 мин 34 с	24 мин 24 с
Intel Core i7-8750H CPU 2.20GHz	Windows 10	GTX 1050 Ti (4 Гб, notebook)	1280	8 мин 44 с	8 мин 15 с
Intel Core i7-10875H CPU 2.30GHz		RTX 2070 (8 Гб, notebook)	1280	2 мин 27 с	2 мин 50 с
Intel Core i7-10700F CPU 2.90GHz	Linux (Ubuntu 21.10)	RTX 3080 (10 Гб, desktop)	1280	1 мин 30 с	1 мин 53 с

Как следует из табл. 4, при использовании современного графического вычислителя (*RTX 3080*) скорость обработки изображений достигает 11 изображений в секунду, что значительно превышает возможности человека. На более старой конфигурации аппаратной части (*GTX 960M*) скорость обработки составляет одно изображение за 3 с, что сопоставимо со скоростью обработки изображений человеком. Если рассматривать компьютеры на чипе M1 при условии, что размер изображения на инференсе составляет 640 пикселей, то скорость обработки составит одно изображение за 5 с.

Подсчет животных с помощью регистраций. Для экологических исследований одним из источников информации для подсчета популяции животных служат фото- и видеоизображения с фотоловушек. При срабатывании фотоловушка делает серию из 1–7 изображений. В [15, 26] во время разметки данных класс присваивается именно последователь-

ности изображений, а не одному фотоизображению. Главная проблема такого подхода в том, что не на всех изображениях может присутствовать животное.

С помощью разработанной системы создается таблица с регистрациями, которые позволяют не только подсчитывать число животных в рамках определенного интервала времени, но и уточнять предсказания нейросетевых алгоритмов. Таблица с регистрациями содержит информацию о числе животных на каждом фотоизображении и информацию об изображениях, принадлежащих одному событию, которое называется «регистрация». У каждого животного есть свое время регистрации. В настоящей работе принято, что у всех время регистрации составляет 30 мин. Схема алгоритма формирования регистраций приведена на рис. 3, пример работы одной регистрации — на рис. 4.



Рис. 3. Схема алгоритма формирования регистраций

Событие фотоловушки состоит из нескольких фотоизображений, которые сделаны в течение 2 мин. Координата времени позволяет объединять фотоизображения в группы, полагая, что на них в одной регистрации находятся одни и те же особи. Система допустила ошибки в подсчете объектов, рассматриваемая в качестве примера нейронная сеть не смогла справиться с ситуациями, когда один барсук загораживал другого, — она интерпрети-

рвала их за один объект (рис. 4, в, з). С помощью регистраций эти предсказания можно скорректировать, учитывая предсказания, сделанные на предыдущих и последующих фотоизображениях (рис. 4, а, б), где барсуки были четко разделимы (параметр *Corrected predicted count*). В данном примере в качестве параметра *Corrected predicted count* выбрано максимальное число предсказанных объектов ($\max(\text{Predicted_count}) = 2$) в рамках одной регистрации. Регистрации также позволяют корректировать предсказания сети не только по числу животных на фотоизображении, но и по предсказанным классам.

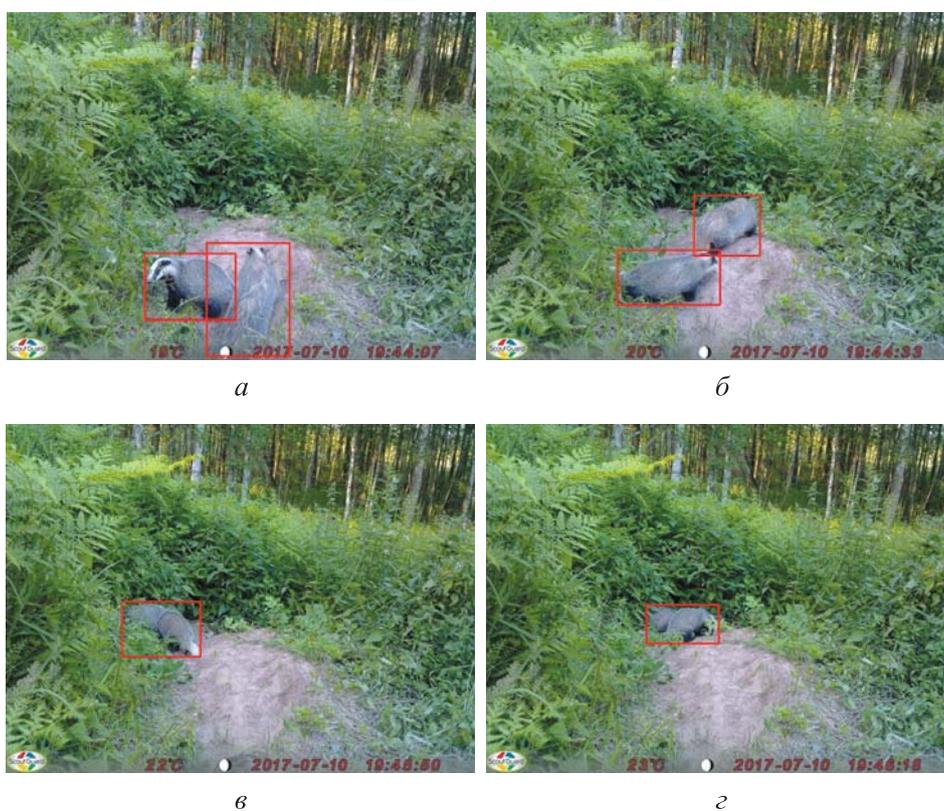


Рис. 4. Пример регистраций:

а-г — истинное *GT count* = 2 и скорректированное с помощью регистраций *Corrected predicted count* = 2 число особей; истинный и предсказанный двухстадийной системой вид животных — барсук; число особей, предсказанное двухстадийной системой *Predicted count* = 2 (*а, б*) и 1 (*в, з*)

Заключение. Приведено описание системы анализа данных для оперативного дистанционного мониторинга природных территорий на основе технологии анализа фото- и видеоизображений, полученных с фотоловушек. Разработанная система позволяет осуществлять оперативный ди-

станционный мониторинг за счет использования двухстадийной нейросетевой обработки данных, полученных с помощью фотоловушек, что упрощает проведение экологического анализа для подсчета популяции конкретного вида животных. Кроме того, обеспечивается возможность дообучения классификатора под видовое разнообразие любого заповедника, что обеспечивает гибкость и масштабируемость системы при использовании в разных природных зонах с различным видовым составом животных. Подсчет числа животных с помощью регистраций позволяет корректировать количественные и качественные предсказания двухстадийной нейросетевой обработки. Проведено сравнение скорости работы системы на различных аппаратных платформах.

В настоящее время разработанная система применяется на территории Центрально-Лесного государственного природного биосферного заповедника. Планируется расширение географии применения и внедрение системы на других особо охраняемых природоохранных территориях. В будущем база детекторов и классификаторов для пользователей может быть расширена за счет выявления архитектур нейронных сетей, наиболее оптимальных с точки зрения качества и скорости их функционирования. Дальнейшее развитие работы возможно путем внедрения современных подходов к обучению нейронных сетей [28], которые позволят значительно сократить время обучения.

Программное обеспечение программно-аппаратного комплекса включено в единый реестр российских программ для электронных вычислительных машин и баз данных (регистрационный номер 160556 от 23.12.2022 г.).

ЛИТЕРАТУРА

- [1] O'Connell A.F., Nichols J.D., Karanth K.U. Camera traps in animal ecology. Berlin, Springer Science & Business Media, 2011.
DOI: <https://doi.org/10.1007/978-4-431-99495-4>
- [2] Желтухина Ю.С., Огурцов С.С., Волкова К.А. и др. Применение фотоловушек в экологическом просвещении школьников и студентов в Центрально-Лесном заповеднике. *Биологическое и экологическое образование студентов и школьников: актуальные проблемы и пути их решения. Матер. IV Междунар. науч.-практ. конф.* Самара, СГСПУ, 2018, с. 98–110.
- [3] Огурцов С.С., Волков В.П., Желтухин А.С. Обзор современных способов хранения, обработки и анализа данных с фотоловушек в зоологических исследованиях. *Nature Conservation Research. Заповедная наука*, 2017, № 2, с. 73–98.
- [4] Алпеев М.А., Артаев О.Н., Варгот Е.В. и др. Первый опыт применения фотоловушек в Мордовском государственном природном заповеднике

имени П.Г. Смидовича. *Труды Мордовского государственного природного заповедника имени П.Г. Смидовича*, 2018, № 20, с. 3–14.

[5] Гаврилов Д.А., Щелкунов Н.Н. Программное обеспечение разметки крупноформатных аэрокосмических изображений и подготовки обучающих выборок. *Научное приборостроение*, 2020, т. 30, № 2, с. 67–75.

DOI: <https://doi.org/10.18358/np-30-2-i6775>

[6] Гаврилов Д.А. Исследование применимости сверточной нейронной сети U-Net к задаче сегментации изображений авиационной техники. *Компьютерная оптика*, 2021, т. 45, № 4, с. 575–579. DOI: <https://doi.org/10.18287/2412-6179-CO-804>

[7] Леус А.В., Ефремов В.А. Применение методов компьютерного зрения для анализа изображений, собранных с фотоловушек в рамках программно-аппаратного комплекса мониторинга состояния окружающей среды на особо охраняемых природных территориях. *Труды Мордовского государственного природного заповедника имени П. Г. Смидовича*, 2021, № 28, с. 121–129.

[8] Chen G., Han T.X., He Z., et al. Deep convolutional neural network based species recognition for wild animal monitoring. *IEEE ICIP*, 2014, pp. 858–862.

DOI: <https://doi.org/10.1109/ICIP.2014.7025172>

[9] Yousif H., Yuan J., Kays R., et al. Animal Scanner: software for classifying humans, animals, and empty frames in camera trap images. *Ecol. Evol.*, 2019, vol. 9, no. 4, pp. 1578–1589. DOI: <https://doi.org/10.1002/ece3.4747>

[10] Yousif H., Yuan J., Kays R., et al. Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. *IEEE ISCAS*, 2017. DOI: <https://doi.org/10.1109/ISCAS.2017.8050762>

[11] He K., Zhang X., Ren S., et al. Deep residual learning for image recognition. *IEEE CVPR*, 2016, pp. 770–778. DOI: <https://doi.org/10.1109/CVPR.2016.90>

[12] Gomez-Villa A., Salazar A., Vargas F. Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecol. Inform.*, 2017, vol. 41, pp. 24–32.

DOI: <https://doi.org/10.1016/j.ecoinf.2017.07.004>

[13] Norouzzadeh M.S., Nguyen A., Kosmala M., et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *PNAS*, 2018, vol. 115, no. 25, pp. 5716–5725. DOI: <https://doi.org/10.1073/pnas.1719367115>

[14] Yu X., Wang J., Kays R., et al. Automated identification of animal species in camera trap images. *EURASIP J. Image Video Proc.*, 2013, vol. 2013, no. 1, art. 52.

DOI: <https://doi.org/10.1186/1687-5281-2013-52>

[15] Tabak M.A., Norouzzadeh M.S., Wolfson D.W., et al. Machine learning to classify animal species in camera trap images: applications in ecology. *Methods Ecol. Evol.*, 2010, vol. 10, no. 4, pp. 585–590. DOI: <https://doi.org/10.1111/2041-210X.13120>

[16] Schneider S., Taylor G.W., Kremer S. Deep learning object detection methods for ecological camera trap data. *IEEE CRV*, 2018, pp. 321–328.

DOI: <https://doi.org/10.1109/CRV.2018.00052>

- [17] Beery S., Van Horn G., Perona P. Recognition in terra incognita. In: *ECCV 2018*. Cham, Springer Nature, 2018, pp. 472–489.
DOI: https://doi.org/10.1007/978-3-030-01270-0_28
- [18] Norouzzadeh M.S., Morris D., Beery S., et al. A deep active learning system for species identification and counting in camera trap images. *Methods Ecol. Evol.*, 2021, vol. 12, no. 1, pp. 150–161. DOI: <https://doi.org/10.1111/2041-210X.13504>
- [19] Whytock R.C., Świeżewski J., Zwerts J.A. Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods Ecol. Evol.*, 2021, vol. 12, no. 6, pp. 1080–1092. DOI: <https://doi.org/10.1111/2041-210X.13576>
- [20] Glenn J. YOLOv5 release v6.1. *github.com: веб-сайт*.
URL: <https://github.com/ultralytics/yolov5/releases/tag/v6.1>
(дата обращения: 15.08.2023).
- [21] Wang C., Yeh I., Liao H.M. You only learn one representation: unified network for multiple tasks. *arXiv:2105.04206*. DOI: <https://doi.org/10.48550/arXiv.2105.04206>
- [22] Ge Z., Liu S., Wang F., et al. YOLOX: exceeding YOLO series in 2021. *arXiv:2107.08430*. DOI: <https://doi.org/10.48550/arXiv.2107.08430>
- [23] Lin T.Y., Maire M., Belongie S., et al. Microsoft COCO: common objects in context. In: *ECCV 2014*. Cham, Springer Nature, 2014, pp. 740–755.
DOI: https://doi.org/10.1007/978-3-319-10602-1_48
- [24] Hu J., Shen L., Albanie S., et al. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, vol. 42, no. 8, pp. 2011–2023.
DOI: <https://doi.org/10.1109/TPAMI.2019.2913372>
- [25] Zhang H., Wu C., Zhang Z., et al. ResNeSt: split-attention networks. *arXiv:2004.08955*. DOI: <https://doi.org/10.48550/arXiv.2004.08955>
- [26] Han D., Yun S., Heo B., et al. ReXNet: diminishing representational bottleneck on convolutional neural network. *arXiv:2007.00992*.
DOI: <https://doi.org/10.48550/arXiv.2007.00992>
- [27] Willi M., Pitman R.T., Cardoso A.W., et al. Identifying animal species in camera trap images using deep learning and citizen science. *Methods Ecol. Evol.*, 2019, vol. 10, no. 1, pp. 80–91. DOI: <https://doi.org/10.1111/2041-210X.13099>
- [28] Tan M., Le Q.V. EfficientNetV2: smaller models and faster training. *arXiv:2104.00298*. DOI: <https://doi.org/10.48550/arXiv.2104.00298>
- [29] Tan M., Le Q.V. EfficientNet: rethinking model scaling for convolutional neural networks. *arXiv:1905.11946*. DOI: <https://doi.org/10.48550/arXiv.1905.11946>
- [30] Tan M., Chen B., Pang R., et al. MnasNet: platform-aware neural architecture search for mobile. *IEEE/CVF CVPR*, 2019, pp. 2815–2823.
DOI: <https://doi.org/10.1109/CVPR.2019.00293>
- [31] Szegedy C., Vanhoucke V., Ioffe S., et al. Rethinking the inception architecture for computer vision. *arXiv:1512.00567*. DOI: <https://doi.org/10.48550/arXiv.1512.00567>
- [32] Sibson R. SLINK: an optimally efficient algorithm for the single-link cluster method. *Comput. J.*, 1973, vol. 16, no. 1, pp. 30–34. DOI: <https://doi.org/10.1093/comjnl/16.1.30>

Ефремов Владислав Александрович — аспирант, программист лаборатории цифровых систем специального назначения МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Леус Андрей Владимирович — канд. техн. наук, ведущий программист лаборатории цифровых систем специального назначения МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Гаврилов Дмитрий Александрович — д-р техн. наук, директор физтех школы ФРТК, МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Мангазеев Даниил Игоревич — магистр, программист лаборатории цифровых систем специального назначения МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Холодняк Иван Витальевич — магистр кафедры радиоэлектроники и прикладной информатики МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Радыш Александра Сергеевна — магистр кафедры радиоэлектроники и прикладной информатики МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Зуев Виктор Александрович — магистр кафедры радиоэлектроники и прикладной информатики МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Водичев Никита Алексеевич — аспирант кафедры радиоэлектроники и прикладной информатики МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Паршиков Михаил Михайлович — магистр кафедры радиоэлектроники и прикладной информатики МФТИ (Российская Федерация, 141701, Московская обл., г. Долгопрудный, Институтский переулок, д. 9).

Просьба ссылаться на эту статью следующим образом:

Ефремов В.А., Леус А.В., Гаврилов Д.А. и др. Система анализа данных с фотоловушек для оперативного дистанционного мониторинга природных территорий. *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение*, 2023, № 4 (145), с. 85–109. DOI: <https://doi.org/10.18698/0236-3933-2023-4-85-109>

CAMERA TRAP DATA ANALYSIS SYSTEM FOR OPERATIONAL REMOTE MONITORING OF THE NATURAL AREAS

V.A. Efremov

A.V. Leus

D.A. Gavrilov

D.I. Mangazeev

I.V. Kholodnyak

A.S. Radysh

V.A. Zuev

N.A. Vodichev

M.M. Parshikov

efremov.va@phystech.edu

leus.av@mipt.ru

gavrilov.da@mipt.ru

mangazeev.di@phystech.edu

kholodnyak.iv@phystech.edu

radysh.as@phystech.edu

zuev.va@phystech.edu

vodichev.na@phystech.edu

parchikov.mm@phystech.edu

MIPT, Dolgoprudnyi, Moscow Region, Russian Federation

Abstract

The paper presents a system to analyze data from camera traps for operational remote monitoring of the natural areas based on the two-stage neural network image processing system containing the server and user components. The server component is designed to process a large amount of data received from different reserves to learn the neural network algorithms. The user component is required for installation in the local reserve computer. The developed system makes it possible to significantly reduce the data processing time from camera traps and simplifies the ecological analysis. Ability to relearn the classifier for species diversity in any reserve without the detector relearning improves recognition quality of the animal species within one specially protected natural area, which makes the system more flexible and scalable. To adjust the algorithm quantitative and qualitative predictions, software is supplemented with a functionality that makes it possible to automatically create the so-called registrations. Registrations are used to count the number of objects in each photograph taking into account contextual information from the image sequence. Registrations allow adjusting the neural network predictions not only by the number of animals in the photo, but also by the predicted classes. System speeds were compared on various hardware platforms. It is shown that introduction of the advanced graphic computers makes it possible to process images at a rate significantly exceeding human capabilities

Keywords

Camera trap images, agglomerative clustering, deep convolutional neural networks, detection, classification, two-stage approach, registrations

Received 21.02.2023

Accepted 27.03.2023

© Author(s), 2023

REFERENCES

- [1] O'Connell A.F., Nichols J.D., Karanth K.U. Camera traps in animal ecology. Berlin, Springer Science & Business Media, 2011.
DOI: <https://doi.org/10.1007/978-4-431-99495-4>
- [2] Zheltukhina Yu.S., Ogurtsov S.S., Volkova K.A., et al. [The use of camera traps in the environmental education of school children and students in the central forest reserve]. *Biologicheskoe i ekologicheskoe obrazovanie studentov i shkolnikov: aktualnye problemy i puti ikh resheniya. Mater. IV Mezhdunar. nauch.-prakt. konf.* [Biological and Ecological Education of Students and Schoolchildren: Current Problems and Ways to Solve Them. Proc. IV Int. Sc.-Pract. Conf.]. Samara, SGSPU Publ., 2018, pp. 98–110 (in Russ.).
- [3] Ogurtsov S.S., Volkov V.P., Zheltukhin A.S. Review of some actual methods of storage, processing and analysis of data from camera traps in zoological research. *Nature Conservation Research. Zapovednaya nauka* [Nature Conservation Research], 2017, no. 2, pp. 73–98 (in Russ.).
- [4] Alpeev M.A., Artaev O.N., Vargot E.V., et al. The first results of the camera trap use in the Mordovia State Nature Reserve. *Trudy Mordovskogo gosudarstvennogo prirodnogo zapovednika imeni P.G. Smidovicha* [Proceedings of the Mordovia State Nature Reserve], 2018, no. 20, pp. 3–14 (in Russ.).
- [5] Gavrilov D.A., Shchelkunov N.N. Software for large format aerospace image marking and training samples preparation. *Nauchnoe priborostroenie*, 2020, vol. 30, no. 2, pp. 67–75 (in Russ.). DOI: <https://doi.org/10.18358/np-30-2-i6775>
- [6] Gavrilov D.A. Investigation of the applicability of the convolutional neural network U-Net to a problem of segmentation of aircraft images. *Kompyuternaya optika* [Computer Optics], 2021, vol. 45, no. 4, pp. 575–579 (in Russ.).
DOI: <https://doi.org/10.18287/2412-6179-CO-804>
- [7] Leus A.V., Efremov V.A. Computer vision methods application for camera traps image analysis within the software for the reserves environmental state monitoring. *Trudy Mordovskogo gosudarstvennogo prirodnogo zapovednika imeni P.G. Smidovicha* [Proceedings of the Mordovia State Nature Reserve], 2021, no. 28, pp. 121–129 (in Russ.).
- [8] Chen G., Han T.X., He Z., et al. Deep convolutional neural network based species recognition for wild animal monitoring. *IEEE ICIP*, 2014, pp. 858–862.
DOI: <https://doi.org/10.1109/ICIP.2014.7025172>
- [9] Yousif H., Yuan J., Kays R., et al. Animal Scanner: software for classifying humans, animals, and empty frames in camera trap images. *Ecol. Evol.*, 2019, vol. 9, no. 4, pp. 1578–1589. DOI: <https://doi.org/10.1002/ece3.4747>
- [10] Yousif H., Yuan J., Kays R., et al. Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. *IEEE ISCAS*, 2017. DOI: <https://doi.org/10.1109/ISCAS.2017.8050762>

- [11] He K., Zhang X., Ren S., et al. Deep residual learning for image recognition. *IEEE CVPR*, 2016, pp. 770–778. DOI: <https://doi.org/10.1109/CVPR.2016.90>
- [12] Gomez-Villa A., Salazar A., Vargas F. Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecol. Inform.*, 2017, vol. 41, pp. 24–32. DOI: <https://doi.org/10.1016/j.ecoinf.2017.07.004>
- [13] Norouzzadeh M.S., Nguyen A., Kosmala M., et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *PNAS*, 2018, vol. 115, no. 25, pp. 5716–5725. DOI: <https://doi.org/10.1073/pnas.1719367115>
- [14] Yu X., Wang J., Kays R., et al. Automated identification of animal species in camera trap images. *EURASIP J. Image Video Proc.*, 2013, vol. 2013, no. 1, art. 52. DOI: <https://doi.org/10.1186/1687-5281-2013-52>
- [15] Tabak M.A., Norouzzadeh M.S., Wolfson D.W., et al. Machine learning to classify animal species in camera trap images: applications in ecology. *Methods Ecol. Evol.*, 2010, vol. 10, no. 4, pp. 585–590. DOI: <https://doi.org/10.1111/2041-210X.13120>
- [16] Schneider S., Taylor G.W., Kremer S. Deep learning object detection methods for ecological camera trap data. *IEEE CRV*, 2018, pp. 321–328. DOI: <https://doi.org/10.1109/CRV.2018.00052>
- [17] Beery S., Van Horn G., Perona P. Recognition in terra incognita. In: *ECCV 2018*. Cham, Springer Nature, 2018, pp. 472–489. DOI: https://doi.org/10.1007/978-3-030-01270-0_28
- [18] Norouzzadeh M.S., Morris D., Beery S., et al. A deep active learning system for species identification and counting in camera trap images. *Methods Ecol. Evol.*, 2021, vol. 12, no. 1, pp. 150–161. DOI: <https://doi.org/10.1111/2041-210X.13504>
- [19] Whytock R.C., Świeżewski J., Zwerts J.A. Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods Ecol. Evol.*, 2021, vol. 12, no. 6, pp. 1080–1092. DOI: <https://doi.org/10.1111/2041-210X.13576>
- [20] Glenn J. YOLOv5 release v6.1. *github.com: website*. Available at: <https://github.com/ultralytics/yolov5/releases/tag/v6.1> (accessed: 15.08.2023).
- [21] Wang C., Yeh I., Liao H.M. You only learn one representation: unified network for multiple tasks. *arXiv:2105.04206*. DOI: <https://doi.org/10.48550/arXiv.2105.04206>
- [22] Ge Z., Liu S., Wang F., et al. YOLOX: exceeding YOLO series in 2021. *arXiv:2107.08430*. DOI: <https://doi.org/10.48550/arXiv.2107.08430>
- [23] Lin T.Y., Maire M., Belongie S., et al. Microsoft COCO: common objects in context. In: *ESSV 2014*. Cham, Springer Nature, 2014, pp. 740–755. DOI: https://doi.org/10.1007/978-3-319-10602-1_48
- [24] Hu J., Shen L., Albanie S., et al. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, vol. 42, no. 8, pp. 2011–2023. DOI: <https://doi.org/10.1109/TPAMI.2019.2913372>

- [25] Zhang H., Wu C., Zhang Z., et al. ResNeSt: split-attention networks. *arXiv:2004.08955*. DOI: <https://doi.org/10.48550/arXiv.2004.08955>
- [26] Han D., Yun S., Heo B., et al. ReXNet: diminishing representational bottleneck on convolutional neural network. *arXiv:2007.00992*. DOI: <https://doi.org/10.48550/arXiv.2007.00992>
- [27] Willi M., Pitman R.T., Cardoso A.W., et al. Identifying animal species in camera trap images using deep learning and citizen science. *Methods Ecol. Evol.*, 2019, vol. 10, no. 1, pp. 80–91. DOI: <https://doi.org/10.1111/2041-210X.13099>
- [28] Tan M., Le Q.V. EfficientNetV2: smaller models and faster training. *arXiv:2104.00298*. DOI: <https://doi.org/10.48550/arXiv.2104.00298>
- [29] Tan M., Le Q.V. EfficientNet: rethinking model scaling for convolutional neural networks. *arXiv:1905.11946*. DOI: <https://doi.org/10.48550/arXiv.1905.11946>
- [30] Tan M., Chen B., Pang R., et al. MnasNet: platform-aware neural architecture search for mobile. *IEEE/CVF CVPR*, 2019, pp. 2815–2823. DOI: <https://doi.org/10.1109/CVPR.2019.00293>
- [31] Szegedy C., Vanhoucke V., Ioffe S., et al. Rethinking the inception architecture for computer vision. *arXiv:1512.00567*. DOI: <https://doi.org/10.48550/arXiv.1512.00567>
- [32] Sibson R. SLINK: an optimally efficient algorithm for the single-link cluster method. *Comput. J.*, 1973, vol. 16, no. 1, pp. 30–34. DOI: <https://doi.org/10.1093/comjnl/16.1.30>

Efremov V.A. — Post-Graduate Student, Programmer, Laboratory of Special Purpose Digital Systems, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Leus A.V. — Cand. Sc. (Eng.), Leading Programmer, Laboratory of Special Purpose Digital Systems, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Gavrilov D.A. — Dr. Sc. (Eng.), Director, Physical and Technical School, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Mangazeev D.I. — Master, Programmer, Laboratory of Special Purpose Digital Systems, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Kholodnyak I.V. — Master, Department of Radioelectronics and Applied Informatics, MIPT (Institutskiy pereulok 9, Dolgoprudny, Moscow Region, 141701 Russian Federation).

Radysh A.S. — Master, Department of Radioelectronics and Applied Informatics, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Zuev V.A. — Master, Department of Radioelectronics and Applied Informatics, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Vodichev N.A. — Post-Graduate Student, Department of Radioelectronics and Applied Informatics, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Parshikov M.M. — Master, Department of Radioelectronics and Applied Informatics, MIPT (Institutskiy pereulok 9, Dolgoprudnyi, Moscow Region, 141701 Russian Federation).

Please cite this article in English as:

Efremov V.A., Leus A.V., Gavrilov D.A., et al. Camera trap data analysis system for operational remote monitoring of the natural areas. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2023, no. 4 (145), pp. 85–109 (in Russ.). DOI: <https://doi.org/10.18698/0236-3933-2023-4-85-109>