

# ПРИМЕНЕНИЕ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ АНАЛИЗА ИЗОБРАЖЕНИЙ, СОБРАННЫХ С ФОТОЛОВУШЕК В РАМКАХ ПРОГРАММНО-АППАРАТНОГО КОМПЛЕКСА МОНИТОРИНГА СОСТОЯНИЯ ОКРУЖАЮЩЕЙ СРЕДЫ НА ОСОБО ОХРАНЯЕМЫХ ПРИРОДНЫХ ТЕРРИТОРИЯХ

**А.В. Леус, В.А. Ефремов**

*Московский физико-технический институт  
(Национальный исследовательский университет), Россия  
e-mail: leus.av@phystech.edu, efremov.va@phystech.edu*

В работе рассматривается применение методов компьютерного зрения для автоматической обработки фото- и видеоданных, собранных с фотоловушек. Предлагается двухстадийный подход обработки данных, состоящий из детектирования объектов на фотографиях и их последующей классификации. Разработано программное обеспечение, состоящее из серверной и пользовательской частей. Серверная часть позволяет обучать детектор и классификатор. Пользовательская часть состоит из двух модулей, первый из которых позволяет налаживать процесс дообучения/обучения нейронных сетей для классификации объектов под конкретный ООПТ, а второй – автоматически обрабатывать фото- и видеоданные, поступающие с фотоловушек, при помощи обученных нейронных сетей. Эффективность работы программы демонстрируется качеством распознавания животных с точностью выше 95%.

**Ключевые слова:** фотоловушка, разметка данных, обучающая и контрольная выборки, детекция, классификация, программное обеспечение, IOU

## Введение

В настоящее время для сбора фото- и видеоматериалов в заповедниках распространены фотоловушки. Материалы позволяют осуществлять мониторинг территорий, решать различные задачи по определению ареалов обитания животных, повышать безопасность на территориях ООПТ. Вместе с тем, для эффективной работы с фотографиями и видеофайлами требуются значительные трудозатраты, связанные с сортировкой, классификацией и анализом изображений. Сейчас большую часть этой тяжелой работы выполняют сотрудники заповедников, министерства и отдельные волонтеры, что приводит к избыточному расходу как материальных, так и интеллектуальных ресурсов. Кроме того, усталость и ограниченная скорость обработки данных оператором может привести к ошибкам и не позволяет обработать большой объем данных, собранных даже в рамках одного заповедника, за ограниченное время. Если ставится задача собрать и оперативно обработать данные с нескольких заповедников или со всех заповедников страны, то без сложной системы автоматизации решить указанную задачу невозможно. В связи с этим коллективом лаборатории цифровых систем специального назначения МФТИ при поддержке научных сотрудников, работающих в заповедниках, экспертов Министерства природных ресурсов и экологии Российской Федерации был разработан программный комплекс (Гаврилов и др., 2018, 2021) была сформулирована задача автоматизировать процесс сортировки и анализа фото- и видеоданных фотоловушек.

Данные о дате и времени съемки фотографии видео сохраняются в файле при его создании внутри фотоловушки. Остается выделить объект на изображении и верно его классифицировать. Под объектами было принято определять изображенных животных, классифицируя их до вида (далее будем вид животного также называть классом), отмечать на изображениях людей (без идентификации) и различную используемую в заповедниках транспортную технику (без дополнительной классификации). Данные о месте съемки определяются местоположением фотоловушки. Таким образом, данные с каждой фотоловушки маркировались или помещались в отдельную папку и после обработки однозначно определяли место

регистрации объектов. Что в совокупности позволяет осуществлять мониторинг с учетом, что известно время, место и класс (вид животного, человек или техника) зарегистрированных с помощью фотоловушек объектов.

В итоге, наиболее сложной и ресурсоемкой подзадачей оказалась задача детектирования и классификация объектов. Для ее решения необходимо было собрать достаточно обширные выборки фото и видео данных, разметить их под детекцию и классификацию объектов и создать нейронную сеть, способную с высокой вероятностью классифицировать исходные данные.

### **Сбор и описание данных**

Для формирования выборки более 50 заповедников предоставили необработанные или частично обработанные данные с фотоловушек. Общий объем фотографий и видеофайлов превысил 3000 Гигабайт. Часть данных была также загружена из различных сервисов фотографий сети Интернет. В общей сложности была сформирована выборка из 200 тыс. фотографий, размеченных для детекции объектов и 400 тыс. фотографий, размеченных для классификации объектов на 25 классов.

Наилучшие результаты при обучении и анализе были получены на данных, к которым предъявляются следующие требования:

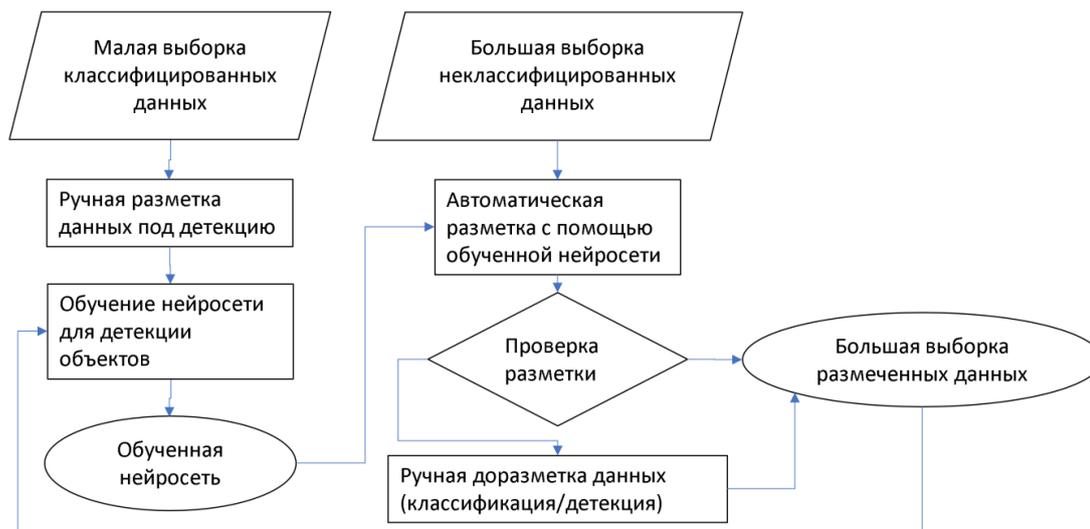
- разрешение фотографий не должно быть ниже  $720 \times 1280$  пикселей и не более 40 мегапикселей;
- в контрольной выборке объекты должны быть изображены не менее чем на 50% фотографий выборки;
- изображение объекта на исходной фотографии должно быть размером не менее  $50 \times 50$  пикселей и занимать не более 80% площади всего изображения;
- объект на изображении не должен быть скрыт более чем на 50% другими объектами, животными или краем поля зрения камеры фотоловушки;
- контрольная разметка должна быть не противоречива, т.е. могут быть использованы только те фотографии, на которых 3 эксперта независимо детектировали и одинаково классифицировали одно то же животное.

Разметка под детекцию (рис. 1) подразумевает выделение объектов на каждой фотографии специальной ограничивающей рамкой (Bounding box – BBox). Разметка под классификацию была сформирована как набор фотографий, преимущественно квадратных, одного класса, сохраненных в папку с названием этого класса. Причем характерный размер объекта соответствует размеру фотографии.

Для создания выборки использовалась следующая схема:

- для каждого класса была подготовлена малая выборка из 1000 фотографий для детекции на каждый класс;
- на основании этой выборки была обучена детектирующая и классифицирующая однопроходная нейронная сеть на базе архитектуры YoloV5 (Redmon, Farhadi, 2018);
- с помощью обученной нейронной сети была произведена автоматическая разметка большего объема данных, которые затем были проверены и при необходимости переразмечены вручную;
- на основе новой большой выборки размеченных данных осуществлялось новое обучение нейронной сети и весь процесс итеративно повторялся.

После первичного обучения полученные нейросетевые алгоритмы передаются на пункт обработки данных и разметки. При последующей обработке данных первичную обработку осуществляют с помощью нейросети. Если входной объем обрабатываемых данных небольшой (до 10000 фотоизображений или до 500 видеофайлов), то обработка запускается непосредственно на пункте обработки данных (Gavrilov, 2019). Если необходимо обработать данные большего объема, то они копируются на сервер, после чего данные автоматической разметки передаются обратно на пункт обработки данных. После первичной нейросетевой обработки разметчик проверяет данные и исправляет их, формируя новую обучающую



**Рис. 1.** Общая схема разметки.  
**Fig. 1.** General markup scheme.

выборку. Когда количество проверенных разметчиком изображений достигает 10000, считается, что сформирована обучающая выборка на один класс.

### Описание системы, структурная схема и алгоритмы

Учитывая, что датчик движения фотоловушки в основном реагирует на теплокровных и достаточно крупных животных, далее исходим из того, что на территории российских ООПТ фотоловушками на регулярной основе фиксируются десятки видов животных. При этом чем крупнее животное, тем более часто срабатывает детектор движения, регистрируя появление животного перед камерой. Для мелких животных, скрытых под листвой, в траве или за деревьями вероятность сработки фотоловушки минимальна. В рамках данного исследования в качестве самого мелкого животного была выбрана белка, но в процессе постобработки данных в рамках мониторинга необходимо учитывать, что датчик движения фотоловушки может не всегда срабатывать на движение животного данного размера (Lovtsov, Gavrilov, 2019; Ловцов, Гаврилов, 2019).

Таким образом, может быть сформирован полный список всех видов животных, которые могут быть найдены на фотографиях с фотоловушек. Для автоматизации процесса поиска с помощью сформированной выборки может быть обучен детектор, который находит на картинке животное и классифицирует его. Если обучить единый классифицирующий детектор для всех заповедников, то неизбежно будут проявляться ошибки классификации животных, связанные с большими данными. Так, если в рамках конкретной ООПТ фотоловушка фиксирует 16 видов животных, а нейросеть обучена для обнаружения 40 видов, то даже если ошибок классификации не более одного процента, то нейросеть будет периодически выдавать названия видов животных, входящих в 24 непредставленных в рамках данной ООПТ класса. Кроме того, могут быть очень похожие виды, которые на сегодняшний день сложно различить с помощью нейросетевых методов. В связи с этим было принято решение разделить задачи детектирования и классификации.

Так модуль детектирования объектов на фотографии обучается на всех доступных данных, размеченных для детектирования. После обучения, данный модуль способен выделять на фотографии с фотоловушки изображение животного, человека, техники. На этом этапе не ясно, что это за объект. Известно только, что вероятнее всего он принадлежит к одному из классов. Далее применяется классифицирующий модуль, который обучается отдельно для каждой ООПТ. Он определяет, является ли найденный объект человеком, техникой или одним из представленных в рамках данной ООПТ видов.

В итоге для пользователя разработанная система работает следующим образом. На вход программы подается изображение. Запускается детектор, который находит на изображении один из интересующих нас объектов (животное, человек или техника). Объект выделяется рамкой и вырезается из изображения в виде квадратного изображения так, чтобы он находился ровно в центре квадрата (рис. 2–13). Сторона квадрата выбирается равной максимуму сторон рамки, выделенной детектором. Новое изображение подается на ансамбль классификаторов (Szegedy, 2015; He & Zhang, 2015; Tan & Le, 2021), обученных для той ООПТ, где была снята фотография, (если на исходной фотографии детектором обнаружено несколько объектов, то последовательно вырезается и подается каждый из найденных). Каждый классификатор выдает свою версию, что изображено на данной фотографии. На основании обработки результатов, полученных от ансамбля классификаторов, блок принятия решения выдает итоговый результат, который присваивается как новый атрибут найденного детектором объекта.

Принцип работы пользовательской части СПО представлен на схеме (рис. 2).

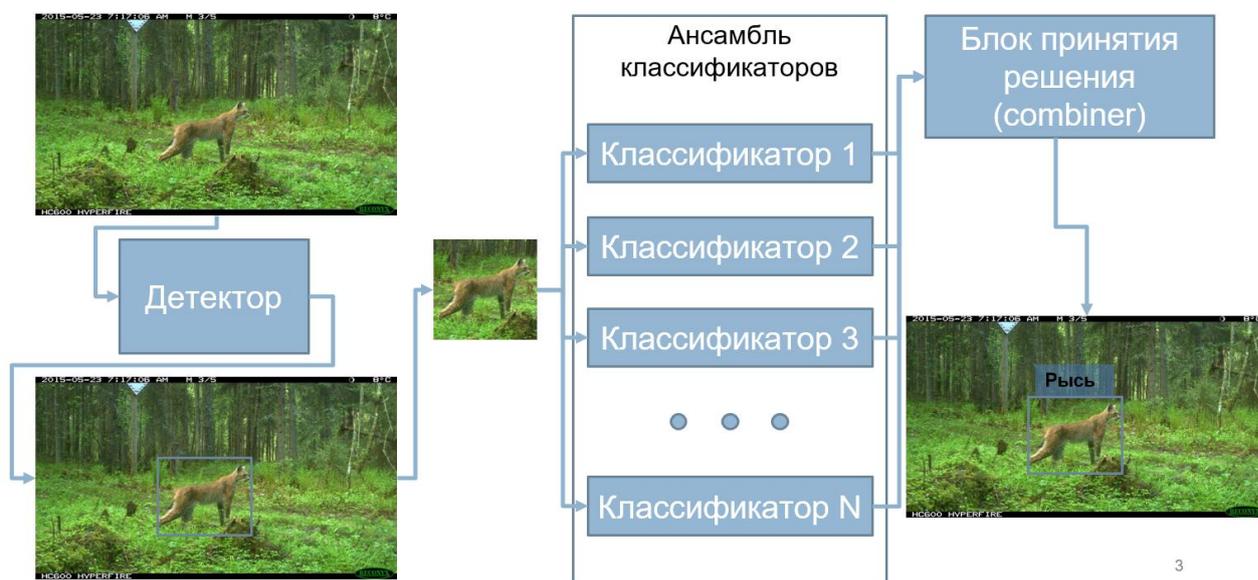


Рис. 2. Принцип работы нейросетевой структуры.

Fig. 2. The principle of operation of a neural network structure.

Серверная часть СПО включает в себя два модуля:

- модуль обучения универсального детектора – модуль, с помощью которого можно по размеченным данным обучить детектор, способный обнаруживать животных, людей и технику (без детальной классификации) на фотографиях;
- модуль обучения классификатора – модуль, с помощью которого по размеченным данным обучить классификатор, способный разделить найденным детектором животных на необходимое количество классов.

Пользовательская часть СПО включает в себя два модуля:

- модуль дообучения / переобучения классификатора, который позволяет настроить универсальный классификатор под непосредственные требования заповедника;
- модуль автоматической обработки данных, который позволяет с помощью универсального детектора (ранее обученного на мощном сервере (Гаврилов, Щелкунов..., 2015)) и адаптированного под местное видовое разнообразие классификатора находить в потоке или архиве фото- и видеоизображений животных и классифицировать их.

Общая схема работы ПО представлена на рисунке 3.



Рис. 3. Общая схема работы ПО.  
 Fig. 3. General scheme of software.

Примеры работы с изображениями животных



Рис. 4. *Alces alces* (лось).  
 Fig. 4. *Alces alces*.



Рис. 5. *Sus scrofa* (кабан).  
 Fig. 5. *Sus scrofa*.



Рис. 6. *Canis lupus* (волк).  
 Fig. 6. *Canis lupus*.



Рис. 7. *Nyctereutes procyonoides* (енотовидная собака).  
 Fig. 7. *Nyctereutes procyonoides*.



Рис. 8. *Vulpes vulpes* (лисица).

Fig. 8. *Vulpes vulpes*.



Рис. 9. *Meles meles* (барсук).

Fig. 10. *Meles meles*.



Рис. 10. *Ursus actos* (бурый медведь).

Fig. 10. *Ursus actos*.



Рис. 11. *Lepus europaeus* / *Lepus timidus* (заяц-русак / заяц-беляк).

Fig. 11. *Lepus europaeus* / *Lepus timidus*.



Рис. 12. *Lynx lynx* (рысь).

Fig. 12. *Lynx lynx*.



Рис. 13. Человек и техника.

Fig. 13. Man and technique.

### Текущие результаты

В рамках исследования был разработан инструмент, позволяющий сотруднику заповедника автоматизировать обработку данных, собранных с фотоловушки. Используемое программное приложение включает в себя уже обученный детектор, способный находить объекты интереса на фотографиях. Также с помощью приложения есть возможность обучить классификатор непосредственно для данной ООПТ. Для этого в интерфейсе приложения (рис. 14) оператор должен выбрать виды животных, представленных в заповеднике.

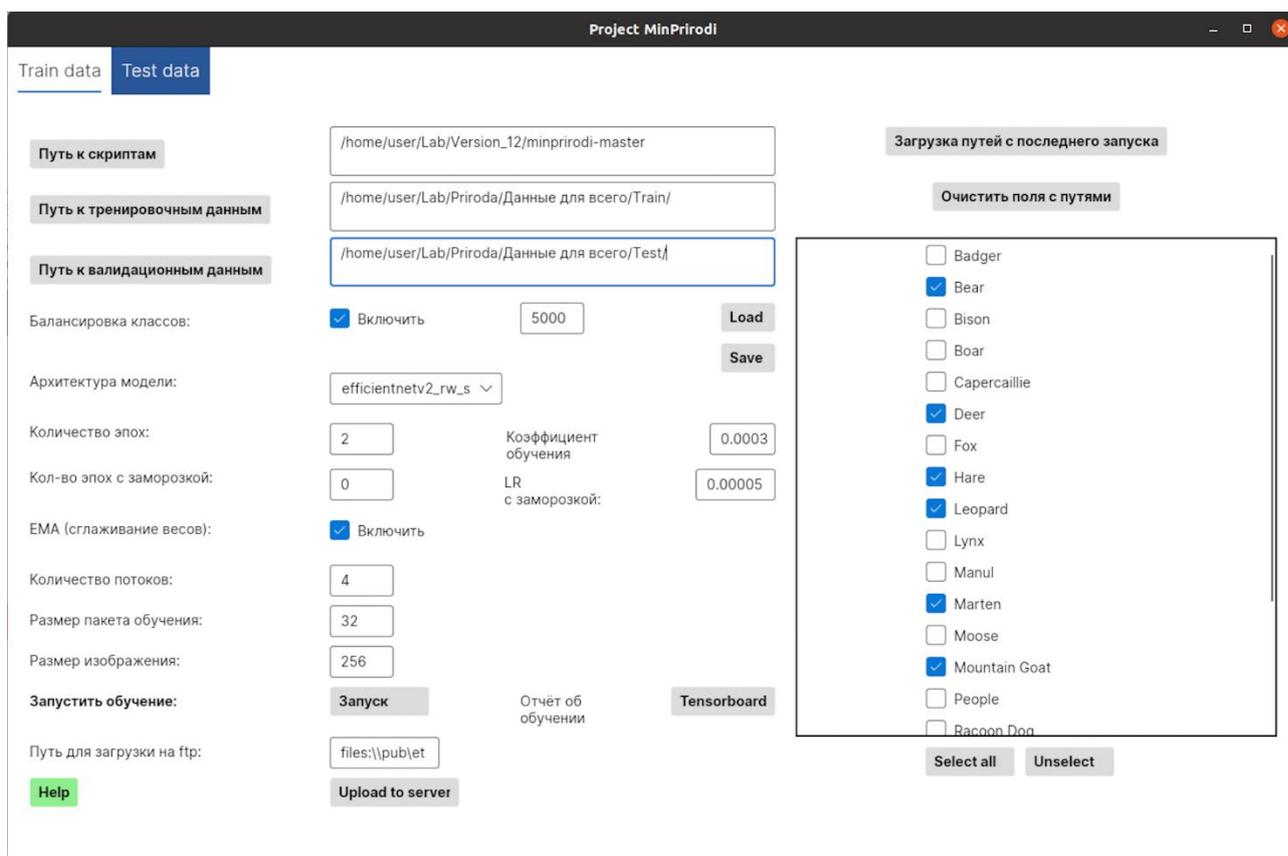


Рис. 14. Интерфейс приложения.

Fig. 14. Software interface

После обучения, которое требует в зависимости от объема данных (He, 2018) и выбранной архитектуры сети различных вычислительных ресурсов, оператор может обрабатывать сырые фото- и видеоданные, для которых формируется достаточно полное описание того, что на них было обнаружено системой в виде таблицы. Дополнительно в качестве отдельной таблицы формируются регистрации, которые позволяют дать количественную оценку обнаруженных животных, а также учитывают контекст сцены (локация, время, последовательность событий).

В настоящий момент для качественных непротиворечивых данных были получены следующие результаты:

– вероятность корректного детектирования объекта более 90%, определяющаяся с помощью Intersection over union (IOU) – метрики степени пересечения между двумя ограничивающими рамками, определяющаяся по формуле (1):

$$IOU = \frac{S(A \cap B)}{S(A \cup B)} > 70\% \quad (1)$$

– вероятность правильного выбора класса объекта для классифицирующего модуля специально подготовленного для конкретного заповедника > 95 %.

Представленные результаты были получены в рамках стендовой отработки данных, собранных с новых фотоловушек в 2021 г., которые не вошли ни в одну из обучающих выборок для создания детектора и классификатора. Испытание проводилось в Государственном природном заповеднике федерального значения «Центрально-Лесной» 17 июня 2021 г.

### Перспективы данной работы

В перспективе планируется расширить число классов так, чтобы в ООПТ можно было обучить собственный классифицирующий модуль, способный различать по видам всех часто

встречающихся на фотоловушках животных. Для этого будет продолжаться формирование соответствующей классифицирующей выборки.

Дополнительно ведется работа по повышению эффективности детектора. Ожидается, что увеличение количества и качества обучающих выборок приведет к существенному росту показателя правильной вероятности обнаружения объектов интереса с помощью детектора до более чем 95 %.

Еще одной задачей в рамках данного исследования рассматривается изучение возможности переноса данной задачи с мощных стационарных вычислительных машин на малые энергоэффективные вычислители, которые могли бы быть размещены в непосредственной близости к фотоловушке и позволили бы оптимизировать процесс глобального автоматического сбора и передачи данных с фотоловушек.

### **Заключение**

В рамках данной работы были найдены подходы к решению наиболее сложной и ресурсоемкой подзадачи автоматизации процесса мониторинга состояния окружающей среды на особо охраняемых природных территориях с помощью применения методов компьютерного зрения для анализа изображений, собранных с фотоловушек. Для ее решения были собраны и размечены выборки фото и видео данных, обучены детектирующие и классифицирующие нейронные сети, способные с достаточно высокой точностью обрабатывать исходные данные.

Эти наработки способны на порядок ускорить процесс обработки сырой информации за счет сокращения до минимума рутинных операций сотрудников заповедника, а также существенного увеличения самой скорости обработки данных.

### **Список литературы**

Гаврилов Д.А., Ивкин А.В., Щелкунов Н.Н. 2018. Система тестирования алгоритмов стабилизации видеоизображений, функционирующих в режиме реального времени // Вестник Московского государственного технического университета им. Н.Э. Баумана. Серия Приборостроение. №6(123). С. 22–36.

Гаврилов Д.А., Щелкунов Н.Н. 2015. Патент на изобретение 2543962 Российская Федерация С2. Аппаратно-вычислительный комплекс виртуализации и управления ресурсами в среде облачных вычислений. 10.03.2015. Заявка № 2013118640/08 от 23.04.2013.; патентообладатель Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Московский физико-технический институт (государственный университет)».

Гаврилов Д.А., Михайлов В.А., Еременко А.Г., Лапушкин А.Г., Ефимов В.Ю., Леус А.В. 2021. Свидетельство о гос. регистрации программы для ЭВМ № 2021663816 от 24.08.2021. Специальное программное обеспечение анализа данных подсистемы наземных пунктов видеонаблюдения.

Ловцов Д.А., Гаврилов Д.А. 2019. Эффективная автоматизированная оптико-электронная система аэрокосмического мониторинга // Правовая информатика. №2. С. 26–39.

Gavrilov D.A. 2019. Quality assessment of objects detection and localization in a video stream // Herald of the Bauman Moscow State Technical University Series Instrument Engineering. №2(125). С. 40–55.

He K., Zhang X. 2015. Deep Residual Learning for Image Recognition // [arXiv:1512.03385](https://arxiv.org/abs/1512.03385)

He T. 2018. Bag of Tricks for Image Classification with Convolutional Neural Networks // [arXiv:1812.01187](https://arxiv.org/abs/1812.01187)

Lovtsov D., Gavrilov D. 2019. Automated special purpose optical electronic system's functional diagnosis // International Seminar on Electron Devices Design and Production, SED 2019 – Proceedings. P. 8798409.

Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. // [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)

Szegedy C. 2015. Rethinking the Inception Architecture for Computer Vision // [arXiv:1512.00567](https://arxiv.org/abs/1512.00567)

Tan M., Le Q. 2021. EfficientNetV2: Smaller Models and Faster Training // [arXiv:2104.00298](https://arxiv.org/abs/2104.00298)

References

- Gavrilov D.A., Mikhailov V.A., Eremenko A.G., Lapushkin A.G., Efimov V.Yu., Leus A.V. 2021. Certificate of State Registration of a computer program No. 2021663816 dated 08.24.2021. Special software for data analysis of the subsystem of ground video surveillance points. [In Russian]
- Gavrilov D.A. 2019. Quality assessment of objects detection and localization in a video stream // Herald of the Bauman Moscow State Technical University Series Instrument Engineering. No. 2(125). P. 40–55.
- Gavrilov D.A., Ivkin A.V., Shchelkunov N.N. 2018. Testing system for video stabilization algorithms operating in real time // Bulletin of the Moscow State Technical University N.E. Bauman. Instrument making series. No. 6(123). P. 22–36. [In Russian]
- Gavrilov D.A., Shchelkunov N.N. 2015. Patent for invention 2543962 Russian Federation C2. Hardware-computing complex for virtualization and resource management in the cloud computing environment. 03.10.2015. Application No. 2013118640/08 dated 23.04.2013.; patentee Federal State Autonomous Educational Institution of Higher Professional Education "Moscow Institute of Physics and Technology (State University)». [In Russian]
- He K., Zhang X. 2015. Deep Residual Learning for Image Recognition. // [arXiv:1512.03385](https://arxiv.org/abs/1512.03385)
- He T. 2018. Bag of Tricks for Image Classification with Convolutional Neural Networks // [arXiv:1812.01187](https://arxiv.org/abs/1812.01187)
- Lovtsov D.A., Gavrilov D.A. 2019. An efficient automated optoelectronic aerospace monitoring system // Legal informatics. No. 2. P. 26–39. [In Russian]
- Lovtsov D., Gavrilov D. 2019. Automated special purpose optical electronic system's functional diagnosis // International Seminar on Electron Devices Design and Production, SED 2019 – Proceedings. P. 8798409.
- Redmon, J, Farhadi A. 2018. YOLOv3: An Incremental Improvement // [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)
- Szegedy C. 2015. Rethinking the Inception Architecture for Computer Vision // [arXiv:1512.00567](https://arxiv.org/abs/1512.00567)
- Tan M., Le Q. 2021. EfficientNetV2: Smaller Models and Faster Training // [arXiv:2104.00298](https://arxiv.org/abs/2104.00298)

**COMPUTER VISION METHODS APPLICATION FOR CAMERA TRAPS IMAGE ANALYSIS WITHIN THE SOFTWARE FOR THE RESERVES ENVIRONMENTAL STATE MONITORING**

**A.V. Leus, V.A. Efremov**

*Moscow Institute of Physics and Technology (National Research University), Russia  
e-mail: leus.av@phystech.edu  
e-mail: efremov.va@phystech.edu*

The paper discusses the use of computer vision methods for camera traps photo and video data automatic processing. A two-stage approach to data processing is proposed, which consists of detecting objects in photos and their subsequent classification. The software consisting of server and user parts has been developed. The server part allows to train the detector and classifier. The user part consists of two modules. The first one allows you to establish the process of training / retraining of neural networks for the classification of objects for a specific reserve. The second allows to process camera traps photo and video data with trained neural networks. The animal recognition program demonstrates the quality of the classification with better than 95% of accuracy.

**Key words:** camera trap, data markup, training and control samples, detection, classification, software, IOU