Tracking Wild Animal Migrations Using Images Captured by Camera Traps Based on Deep Learning Networks

Margarita N. Favorskaya, Andrei I. Pakhirka, Alexander G. Zotin
Reshetnev Siberian State University of Science and Technology,
Krasnoyarsk, Russian Federation

Abstract Tracking of the wild animal displacements in natural parks involves an analysis of large amount of visual data obtained by several camera traps. The problem statement connects with a matching of individual images with other images of the same species. The accumulated volume of video materials allows to solve the problem based on Deep Learning technique, in particular the Siamese neural network. The proposed method was tested using a dataset obtained on the territory of Ergaki natural park, Krasnoyarsk Territory, Russia.

Keywords: monitoring of wild animals, camera traps, image processing, deep learning.

ОТСЛЕЖИВАНИЕ ПЕРЕМЕЩЕНИЙ ДИКИХ ЖИВОТНЫХ ПО ВИЗУАЛЬНЫМ ДАННЫМ ФОТОЛОВУШЕК НА ОСНОВЕ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ

Фаворская М.Н.[(1)], *Пахирка А.И.*[(1)], *Зотин А.Г.*[(1)]

[(1)] ФГБОУ ВО «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева», г. Красноярск

Отслеживание перемещений диких животных в природных парках предполагает анализ большого объема визуальных данных, полученных от нескольких фотоловушек. Ставится задача сопоставления изображений конкретной особи с видеоматериалами данного вида животных. Накопленный объем видеоматериалов позволяет решить задачу с использованием технологии Deep Learning, в частности, на основе Siamese Neural Network. Метод протестирован с использованием набора изображений, полученных на территории природного парка «Ергаки», Красноярский край, Россия.

Ключевые слова: мониторинг диких животных, фотоловушки, обработка изображений, глубокое обучение.

Введение. Использование фотоловушек для мониторинга диких животных становится все более популярным способом неинвазивного наблюдения за дикими животными в их среде обитания. Проблема заключается в обработке больших объемов информации, полученных, например, в течение полугода от нескольких десятков фотоловушек, установленных в национальных парках и на особо охраняемых природных территориях. Выбор мест установки определяется работниками парков в местах троп, водопоев, солонцов и т.д. Ставится задача сопоставления визуальных данных конкретной особи с видеоматериалами, полученными от распределенной сети фотоловушек, с целью нахождения локальных миграций данной особи. Такая задача является одной из завершающих задач мониторинга диких животных при условии, что выполнен предварительный отсев малоинформативных снимков и распознан вид животного. Большой накопленный объем видеоматериалов позволяет решить эту проблему с использованием технологии Deep Learning, в частности с использованием сети глубокого обучения Siamese Neural Network. Метод протестирован с использованием набора данных изображений, полученных на территории природного парка «Ергаки», Красноярский край, Россия.

Постановка задачи. Известно, что каждое животное имеет свой ареал обитания, и поэтому его появление может быть зафиксировано ограниченным количеством близлежащих фотоловушек N_F. В силу удаленности территорий национальных парков фотоловушки не связаны в единую информационную сеть и не могут передавать текущую информацию на сервер. Они способны сохранять серию снимков в имеющемся накопителе после срабатывания датчика движения. Обычно длительность между отснятыми снимками составляет 3-5 с, что позволяет получить несколько изображений особи в разных положениях. Фотоловушки срабатывают в любое время суток и при любых метеорологических условиях, что приводит к появлению плохих по качеству снимков или снимков с отсутствием животного. Следует отметить, что появление человека фиксируется аналогичным образом. Серия снимков $S_{im} = \{Im_1, Im_2, Im_3, \ldots, Im_{Ns}\}$, где IM_i – i-й снимок, Ns – количество снимков в серии, имеет атрибуты дата D_s, время T_s, температура K_s. Примем, что срабатывание фотоловушки инициируется одним визуальным объектом $VO_i \in \{VO_1, VO_2, \ldots, VO_{Na}\}$, где VO_{Na} – количество

классов, включая класс неопознанного визуального объекта. Таким образом, мощность множества полученных снимков IM_T за установленный временной интервал определяется как сумма:

$$\{IM_T\} = \sum_{i=1}^{N_F} count\left(S_{im_i}\right)$$

где *count* (·) – функция подсчета количества снимков в *i*-й серии.

Примем, что из полученного множества снимков удалены неинформативные снимки [7], содержащие отсутствие визуального объекта или значительные артефакты, вызванные условиями съемки. При этом неинформативные снимки могут составлять от 60% до 75% от общего объема видеоматериалов [1]. К оставшемуся множеству снимков IM_S применяется процедура составления модели фона (для конкретной фотоловушки), что позволяет достаточно быстро и точно находить визуальный объект [2]. Отобранное множество снимков содержит подмножества изображений видов IM_A и подмножество изображений человека IM_H:

$$IM_S = \left(IM_{A_1}, IM_{A_2}, \ldots, IM_{A_M}\right) \bigcup IM_H,$$

где M – количество видов (классов) животных.

Авторами разработан метод распознавания видов животных с использованием объединенной сверточной нейронной сети, включающей две структуры VGG16 для распознавания изображений головы и части тела животного и одну структуру VGG19 для распознавания изображения животного в целом [3]. Выходы трех структур объединяются, что позволило получить точность распознавания 80,6% Top-1 и 94,1% Top-5 на сбалансированном наборе данных. Для несбалансированного набора данных (что является характерной ситуацией для данной задачи в силу количества животных на конкретной территории и особенностей их поведения) были получены худшие результаты, а именно 38,7% Top-1 и 54,8% Top-5. Тем не менее, данный подход позволяет разделить подмножества изображений по видам животных. Таким образом, мы получаем подмножество заданного вида IM_{Ai}, которое содержит изображения нескольких особей, относящихся к этому виду. Данная информация является исходной для идентификации конкретной особи и отслеживания ее перемещения с использованием видеоматериалов, полученных от N_F фотоловушек.

Технологии Deep Learning. Технологии глубокого обучения (Deep Learning) позволяют автоматически извлекать абстрактные признаки из исходных («сырых») данных за счет того, что каждый слой нейронов использует операции свертки на перекрывающихся небольших по размеру регионах, поступающих от предыдущих слоев. Причем, на первых сверточных слоях размер таких регионов больше, чем на последующих сверточных слоях сети, так как пространственная область сокращается от слоя к слою и применение большой по размеру маски приведет к пропуску структурных особенностей. Последним слоем глубокой нейронной сети, как правило, является слой softmax, выполняющий роль классификатора. Однако для обучения таких сетей требуется большой объем исходных данных, промаркированный вручную. Отметим, что задача обнаружения и распознавания диких животных удовлетворяет таким условиям и поэтому может быть решена с помощью сверточных нейронных сетей.

Применение технологий глубокого обучения для классификации изображений, полученных от фотоловушек, началось несколько лет назад, и с появлением новых архитектур сверточных нейронных сетей точность распознавания повышалась. Одной из первых была работа [8], в которой ставилась задача автоматической идентификации 20 классов животных

при наличии 20 000 изображений. Была достигнута точность распознавания около 38%, что, тем не менее, превышало точность классификации с применением традиционной технологии на основе «портфеля слов». В [9] было показано, что технологию Deep Learning можно успешно применять на небольших наборах изображений, если требуется отделить птиц от млекопитающих (набор данных из 1572 изображений) и распознать два вида млекопитающих (набор данных из 2597 изображений). Сеть предварительно была обучена на наборе данных ImageNet [10]. Как для обнаружения животных на снимках, так и для распознавания их видов применяются различные сети глубокого обучения, например, AlexNet (8 слоев) [11], NiN (Network in Network) (16 слоев) [12], VGG (Visual Geometry Group, Department of Engineering Science, University of Oxford) (22 слоя) [4], GoogLeNet (32 слоя) [13], а также ResNet (18, 34, 50, 101 и 152 слоя) [14]. Точность детектирования при использовании современных сверточных нейронных сетей высока и может достигать 93,8% Top-1 и 98,8% Top-5 [5], однако следует понимать, что такие высокие значения получаются для набора изображений животных, обитающих на конкретной территории в конкретных климатических условиях и при наличии хорошо обученной сети.

Метод идентификации особи по изображениям. Помимо сверточных нейронных сетей с одним входом в последние годы появились такие архитектуры, которые предполагают подачу двух или даже трех изображений одновременно на две или три ветви сети. Такие сети называются двойными (сиамскими) или тройными. При идентификации конкретной особи из ограниченного набора изображений животных данного вида сиамская структура сети будет полезной.

Архитектура сиамской сети впервые была предложена в 1994 г. для верификации подписи, представленной временным рядом [15]. Позже эта архитектура усилена сверточными нейронными сетями для верификации лиц с использованием техники уменьшения размерности [16]. В последние годы сиамские сверточные сети стали использоваться для повторной идентификации людей, лиц, а также как системы сопровождения объектов.

Архитектура сиамской сети такова, что обучение осуществляется путем подачи на входы двух изображений одновременно. Причем, используются две идентичные параллельные нейронные сети, содержащие одинаковые наборы весовых коэффициентов *W*. Общая структура сиамской нейронной сети представлена на рисунке 1. Однако следует отметить, что имеются другие решения, когда традиционная структура данной сети изменяется путем применения неодинаковых наборов весовых коэффициентов [6].

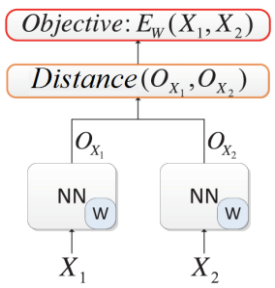

Рис. 1. Общая структура сиамской нейронной сети.

На входы обоих сетей поступают входные изображения X_1 и X_2, их выходами являются векторы признаков O_{X1} и O_{X2}. Далее вычисляется расстояние между этими векторами и

формируется функция $E_W(X_1, X_2)$, которая и характеризует сходство или различие входных изображений. В качестве сетей выбрана архитектура VGG19 (Visual Geometry Group), которая хорошо зарекомендовала себя в задачах распознавания благодаря однотипным конволюционным слоям 3×3 элементов (Conv), слоям подвыборки (Maxpooling) и полносвязным слоям (Fullyconnected) из 4096 нейронов, как показано на рисунке 2.

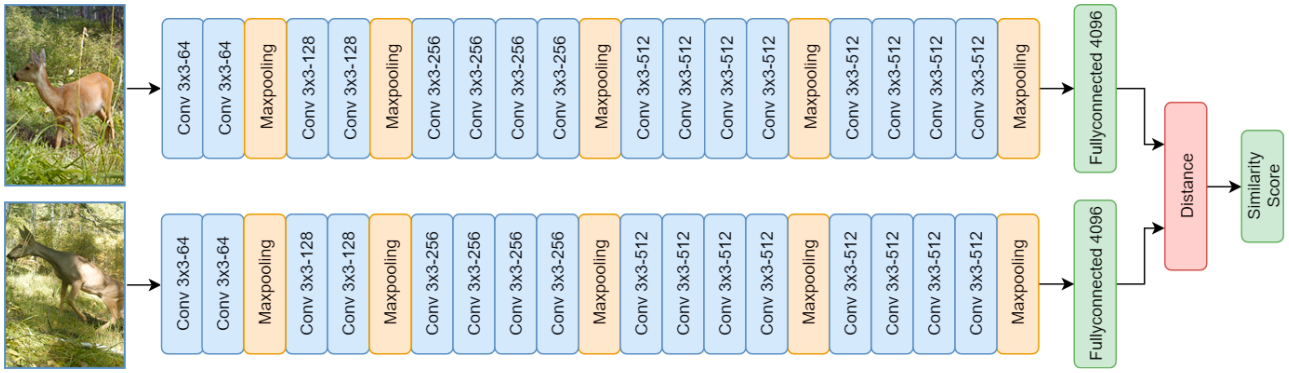

Рис. 2. Структура сиамской нейронной сети на основе VGG19.

Экспериментальные результаты. Для обучения сиамской сверточной сети использовался набор из 842 изображений косуль, полученных на территории природного парка «Ергаки», Красноярский край, Россия. Данный набор содержит изображения 8 особей косули, полученных от 4 фотоловушек, расположенных в разных местах природного парка. На основе исходного набора была сгенерирована обучающая выборка размером порядка 12000 изображений, содержащая случайным образом выбранные пары изображений одной и той же особи, а также разных особей. Далее обучающая выборка была искусственно расширена до 40000 изображений с помощью аугментации (augmentation). Для аугментации применялись следующие действия:

- Поворот изображения на [–45, +45] градусов случайным образом.
- Горизонтальное отражение изображения со случайным сдвигом.
- Масштабирование изображения с коэффициентом [0,75; 1,25].
- Кадрирование изображения.
- Сдвиг каждого цветового канала на [–25,5; +25,5] единиц случайным образом.

Далее полученный набор изображений был разделен на обучающую и тестовую выборки в соотношении 80% на 20%. Сиамская сверточная нейронная сеть была реализована на языке программирования Python 3.5 с использованием фреймворка TensorFlow 1.5. Эксперименты проводились на ПК с процессором Intel i7 3.2GHz, 16GB оперативной памяти, видеокартой 8GB Nvidia GeForce GTX 1070 под операционной системой Windows 7.

Для отслеживания перемещения животного по природному парку требовалось сопоставить изображение конкретной особи с изображениями всех особей данного вида, имеющихся в базе данных. Таким образом, результатом сиамской сверточной нейронной сети являлась оценка подобия двух входных изображений на основе эвклидовой метрики. Результаты сопоставления конкретных особей косули с привязкой к четырем фотоловушкам представлены в таблице.

Таблица. Результаты точности сопоставления особей косули.

	По одному снимку, %	**По серии из 5 снимков, %**
Фотоловушка 1	41,2	56,2
Фотоловушка 2	42,4	58,4
Фотоловушка 3	43,1	57,5
Фотоловушка 4	44,6	59,0

Среднее значение точности сопоставления особей по одному снимку составляет 42,8%, а по серии из пяти снимков – 57,8%. Более точного сопоставления можно добиться увеличением серии тестовых снимков для конкретной особи.

Заключение. Предложенный метод сопоставления визуальных данных конкретной особи с видеоматериалами, полученными от распределенной сети фотоловушек, основан на использовании сиамской сверточной нейронной сети. Предполагается, что изображений особей конкретного вида отобраны, и требуется провести сопоставление снимков для определения миграции особи по природному парку с точностью до мест, в которых установлены фотоловушки. Задача является сложной в силу различных условий съемки и малого количества изображений, пригодных для распознавания конкретной особи. Относительно низкая точность распознавания конкретной особи обусловлена небольшим набором исходных изображений и может быть повышена за счет большей исходной выборки.

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований, Правительства Красноярского края, Красноярского краевого фонда науки в рамках научного проекта (грант № 18-47-240001 а).

ЛИТЕРАТУРА

Статья в журнале:

[1] *Newey S., Davidson P., Nazir S., Fairhurst G., Verdicchio F., Irvine R.J., van der Wal R.* Limitations of recreational camera traps for wildlife management and conservation research: A practitioner's perspective // Ambio. 2015. Vol. 44, pp. 624-635.

[2] *Favorskaya M.N., Buryachenko V.V.* Background extraction method for analysis of natural images captured by camera traps // Informatsionno-upravliaiushchie sistemy [Information and Control Systems]. 2018. Vol. 6, pp. 35-45.

[3] *Favorskaya M., Pakhirka A.* Animal species recognition in the wildlife based on muzzle and shape features using joint CNN. Procedia Computer Science. 2019. (in print).

[4] *Simonyan K., Zisserman A.* Very deep convolutional networks for large-scale image recognition. 2014. arXiv preprint arXiv:1409.1556

[5] *Norouzzadeh M.S., Nguyen A., Kosmala M., Swanson A., Palmer M., Packer C., Clune J.* Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning // PNAS. 2018. Vol. 115, no. 25. pp. E5716-E5725.

[6] *Shaham U., Lederman R.R.* Learning by coincidence: Siamese networks and common variable learning // Pattern Recognition. 2018. Vol. 74. pp. 52-63.

Статья в трудах конференции:

[7] *Favorskaya M.N., Buryachenko V.V.* Selecting informative samples for animal recognition in the wildlife // In: Czarnowski I., Howlett R., Jain L. (eds) Intelligent Decision Technologies 2019. SIST, vol 143, Malta: Springer, 2019. pp. 65-75.

[8] *Chen G., Han T.X., He Z., Kays R., Forrester T.* Deep convolutional neural network based species recognition for wild animal monitoring // Proceedings of the 2014 IEEE International Conference on Image Processing. Paris, France: IEEE, 2014. pp. 858-862.

[9] *Gomez A., Diez G., Salazar A., Diaz A.* Animal identification in low quality camera-trap images using very deep convolutional neural networks and confidence thresholds // Proceedings of the International Symposium on Visual Computing. Las Vegas, Nevada, USA: Springer, 2016. pp. 747-756.

[10] *Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.* Imagenet: A large-scale hierarchical image database // Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA: IEEE, 2009. pp. 248-255.

[11] *Krizhevsky A., Sutskever I., Hinton G.E.* Imagenet classification with deep convolutional neural networks // Proceedings of the 25th International Conference on Neural Information Processing Systems, Vol. 1, Nevads, USA: IEEE, 2012. pp. 1097-1105.

[12] *Lin M., Chen Q., Yan S.* Network in network // Proceedings of the International Conference on Learning Representations, Banff, Canada: IEEE, 2014. pp. 1-10.

[13] *Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A.* Going deeper with convolutions // Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA: IEEE, 2015. pp. 1-9.

[14] *He K. Zhang X., Ren S., Sun J.* Deep residual learning for image recognition // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Las Vegas, NV, USA: IEEE, 2016. pp. 770-778.

[15] *Bromley J., Guyon I., Lecun Y., Säckinger E., Shah R.* Signature verification using a "Siamese" time delay neural network // Proceedings of the Neural Information Processing Systems, Denver, Colorado, USA: IEEE, 1994. pp. 737–744.

[16] *Chopra S., Hadsell R., LeCun Y.* Learning a similarity metric discriminatively, with application to face verification // Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, San Diego, CA, USA: IEEE, 2005. pp. 539–546.