

УДК 004.032.26

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ КЛАССИФИКАЦИИ ДИНАМИЧЕСКИХ ПРИЗНАКОВ ОБЪЕКТОВ

Лаптев Никита Витальевич¹,
nikitalaptev77@gmail.com

Гергет Ольга Михайловна²,
gerget@ipu.ru

Лаптев Владислав Витальевич¹,
vvl39@tpu.ru

Колпащиков Дмитрий Юрьевич¹,
dyk1@tpu.ru

¹ Национальный исследовательский Томский политехнический университет,
Россия, 634050, г. Томск, пр. Ленина, 30.

² Институт проблем управления им. В.А. Трапезникова Российской академии наук,
Россия, 117997, г. Москва, ул. Профсоюзная, 65.

Классификация изображений – классическая задача машинного обучения. Глубокие нейронные сети широко используются в области классификации объектов. Однако проблема анализа объектов с динамически изменяющимися признаками остается актуальной. Для решения этой проблемы авторы предлагают использовать нейронную сеть с долгой краткосрочной памятью. В отличие от классических сверточных нейронных сетей, предлагаемая сеть использует информацию о последовательности изображений, тем самым обеспечивая более высокую точность классификации обнаруженных объектов с динамическими признаками. В исследовании авторы анализируют точность классификации обнаружения дымовых облаков в лесу с использованием различных методов машинного обучения.

Ключевые слова: нейронные сети, традиционное машинное обучение, классификация, изображение, обнаружение пожароопасных ситуаций.

Введение

Количество информации, представленной в виде изображений и видео, растет с каждым годом. Для автоматизации описания такой информации требуются все более сложные алгоритмы и модели извлечения признаков. Одним из наиболее важных направлений при анализе цифровой информации является решение задачи классификации. Методы классификации можно подразделить на две группы: традиционные алгоритмы машинного обучения и искусственные нейронные сети.

Решение задачи классификации объектов на изображениях с помощью технологии традиционного машинного обучения выполняется в два этапа:

1. Выделение признаков. К наиболее распространенным методам извлечения признаков можно отнести дескрипторы функций: HOG [1], LPB [2], SURF [3] и др. На основе полученных в результате обработки признаков формируется вектор признаков.
2. Классификация. Полученные признаки подразделяются на классы с помощью алгоритмов классификации: k-ближайших соседей [4], SVM [5], Random Forest [6] и др.

Стоит отметить, что перечисленные методы извлечения признаков имеют существенные недостатки, связанные с ручным выбором важных, по мнению эксперта, признаков. В следствии чего результат, как правило, зависит от квалификации эксперта. Тем не менее традиционные алгоритмы машинного обучения имеют ряд преимуществ, а именно, низкие требования к вычислительным ресурсам и высокое быстродействие.

Искусственные нейронные сети (ИНС) лишены данных недостатков, но не всегда являются идеальным решением. Как правило, для получения модели ИНС, обладающей высокой обобщающей способностью и хорошей точностью, требуются большие объемы «качественных» данных. Также стоит упомянуть о высоких требованиях к вычислительным ресурсам. Но несмотря на это, популярность использования глубокого обучения полностью оправдана. Нейронные сети автоматизируют этап извлечения признаков: в процессе обучения ИНС стремятся повысить качество формируемого вектора признаков на основе входных данных.

Наибольшей популярностью в решении задачи классификации изображений пользуются сверточные нейронные сети [7]. Несмотря на существование большого количества алгоритмов классификации изображений [8], мало внимания уделяется вопросу классификации видео данных. В случае использования сверточных нейронных сетей для классификации кадров видеопоследовательности необходимо объединить признаки изображений для получения предсказания. Однако при таком подходе будут проигнорированы признаки динамики объектов, поскольку изображения обрабатываются последовательно. Следовательно, данный алгоритм не всегда подходит для классификации видеопоследовательностей.

Рекуррентные нейронные сети [9] имеют более сложную архитектуру и позволяют оценивать динамические признаки, так как способны извлекать временные характеристики. В данной статье авторы предлагают исполь-

зовать рекуррентные нейронные сети с длинной кратковременной памятью [10]. Данный тип нейронных сетей хорошо подходит для решения задач анализа последовательностей. Мы предлагаем архитектуру нейронной сети со слоем Long short-term memory (LSTM) для классификации данных пожароопасных объектов в лесном массиве. Проблема классификации таких данных заключается в отсутствии постоянных признаков у дымового облака, а именно: непостоянность формы, изменение интенсивности цветовой составляющей, непостоянство скорости истечения газа. Анализ видеопоследовательности позволяет выделять динамические признаки, характерные для дымового облака, что дает возможность с более высокой точностью классифицировать искомые на кадре объекты.

Данное исследование является продолжением работы [11]. Основная решаемая задача – повышение качества обнаружения дымового облака в лесном массиве при помощи классификации ранее выделенной области интереса. В этой статье приведен сравнительный анализ применимости различных алгоритмов выделения статических признаков дымового облака, а также представлен анализ сравнения с алгоритмами выделения динамических признаков. Особое внимание уделено разработке модели рекуррентной нейронной сети для выделения динамических признаков в видеопоследовательности.

Постановка задачи

Для решения задачи классификации пожароопасных объектов в лесном массиве было рассмотрено два подхода. Первый подход заключается в анализе одного кадра, взятого из видеопоследовательности посредством технологии машинного обучения. Суть второго подхода состоит в следующем: исходный видеоряд разбивается на последовательность кадров. Из полученной последовательности выбираются кадры в количестве n единиц с равными временными интервалами так, чтобы была возможность проанализировать весь временной интервал видеоряда. Следующим основным этапом является составление вектора признаков для каждого анализируемого кадра и его классификация.

Классификация одного кадра. Составление вектора признаков

Для составления вектора признаков сравниваются следующие модели выделения признаков:

- MobileNetV2;
- ResNet50;
- InceptionV3;
- EfficientNetB1.

Для обучения моделей был собран датасет в размере 8135 изображений. Тестовая выборка составила 2440 изображений, из этих изображений 1220 изображений относились к классу «пожароопасный объект», оставшиеся – «нет пожара». На вход нейронной сети подан тензор фиксированного размера $128 \times 128 \times 3$, количество итераций обучения составило 75. Качественным показателем составления вектора признаков считаем бинарную точность классификации, полученную в результате добавления полносвязного слоя размерностью 1 с линейной функцией активации ReLu (Rectified Linear Unit).

Таблица 1. Результаты моделей нейронных сетей для составления вектора признаков

Table 1. Results of neural network models for compiling a feature vector

Название модели Model name	Вес Weight	Точность Accuracy, %	Время обработки, с Time of processing, s
MobileNetV2	True	0,53	0,053
	False	0,50	0,056
ResNet50	True	0,6	0,074
	False	0,46	0,54
InceptionV3	True	0,62	0,056
	False	0,46	0,05
EfficientNet B1	True	0,6	0,081
	False	0,56	0,054

На основании результатов, приведённых в табл. 1, можно сделать вывод, что наиболее эффективной моделью составления вектора признаков для задачи классификации пожароопасных объектов является модель InceptionV3 с обучаемыми параметрами. Однако точность модели на этом этапе составила всего 62 %.

Классификация на основе вектора признаков

Следующим основным аспектом результирующей модели является классификация на основе вектора признаков. Авторами предлагается провести поиск классификационной модели как среди традиционных методов машинного обучения, так и с использованием построения новой архитектуры ИНС. Для подбора модели классификации среди традиционных алгоритмов машинного обучения воспользуемся сервисом AutoML [12], предлагающим следующие классификационные модели:

- XGBoost [13];
- LightGBM [14];
- Random Forest [15];
- CatBoost [16];
- ElasticNet [17];
- ExtraTree [18].

Также в качестве модели классификации предлагается разработанная авторами архитектура (рис. 1).

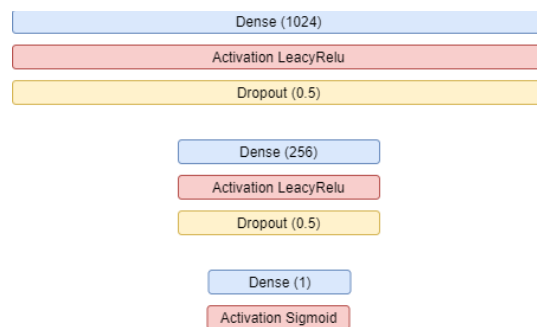


Рис. 1. Архитектура нейронной сети для классификации пожароопасных объектов

Fig. 1. Architecture of the neural network for classification of fire hazardous objects

По результатам, приведенным в табл. 2, предложенная авторами архитектура нейронной сети превосходит по точности классические алгоритмы машинного обучения в задаче классификации дымового облака. Разницу в скорости обработки считаем несущественной. Необходимо от-

метить, что несмотря на достаточно высокие показатели точности, представленные алгоритмы классификации допускают ошибки в идентификации пожароопасных объек-

тов. Данные ошибки могут быть связаны с отсутствием в текущий момент времени в данном кадре дымового облака и отсутствием в анализе динамических признаков.

Таблица 2. Сравнение алгоритмов классификации

Table 2. Comparison of classification algorithms

Показатели моделей Model markers	Традиционные алгоритмы машинного обучения Traditional machine learning algorithms						ИНС/INS
Название модели Model name	XGBoost	LightGBM	Random Forest	CatBoost	ElasticNet	ExtraTree	Собственная архитектура ИНС INS proper architecture
Точность Accuracy	0,82	0,8	0,71	0,5	0,84	0,641	0,836
Время обработки, с Time of processing, s	0,032	0,048	0,057	0,039	0,059	0,044	0,062

Классификация последовательности кадров

Для решения проблемы отсутствия в кадре в текущий момент времени дымового облака, а также для извлечения динамических признаков из исследуемых объектов было рассмотрено два подхода:

- Классификация последовательности изображений с применением среднего арифметического взвешенного результата (рис. 2);

- Классификация последовательности изображений с использованием рекуррентных нейронных сетей (рис. 3).
 Для анализа эффективности классификации видеопоследовательности использовались следующие архитектуры нейронных сетей (рис. 2, 3). Для оценки эффективности классификации использовалась метрика оценки точности и времени обработки кадра. Для обучения моделей использовался фиксированный размер входного тензора $128 \times 128 \times 3$ и последовательность кадров равная 3.

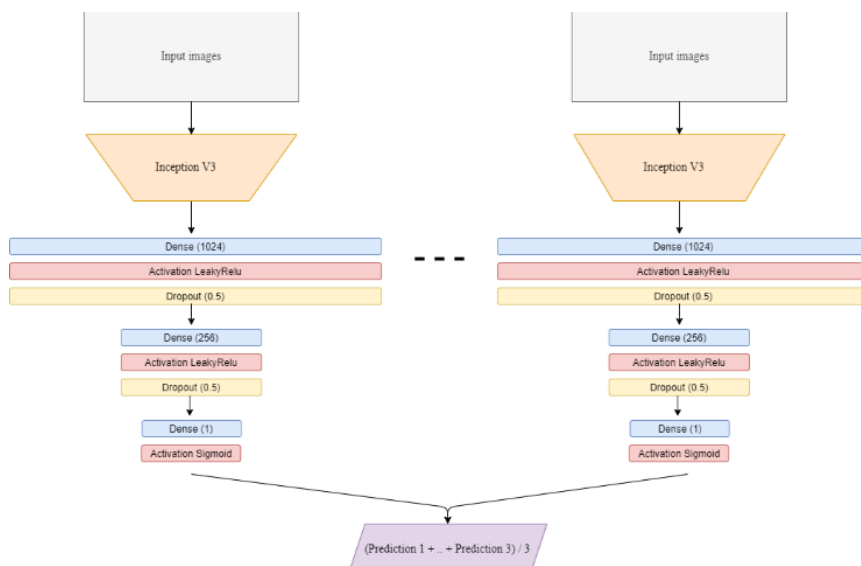


Рис. 2. Архитектура нейронной сети с применением среднего арифметического взвешенного результата

Fig. 2. Architecture of a neural network using the arithmetic mean of the weighted result

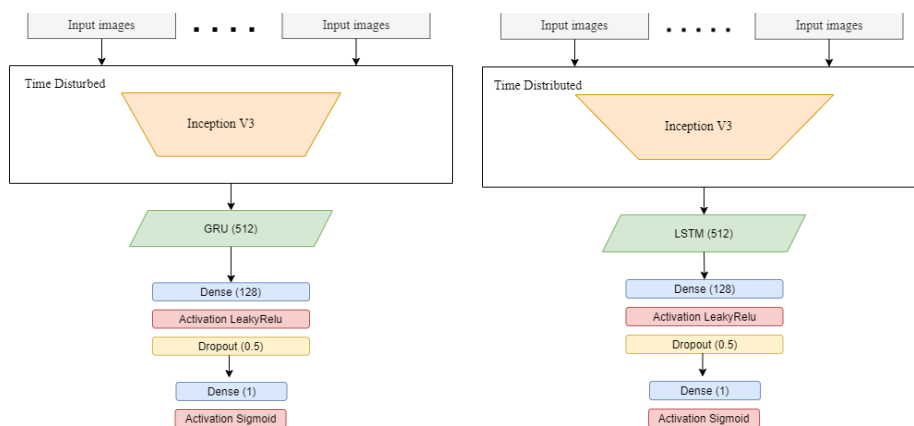


Рис. 3. Архитектуры ИНС

Fig. 3. INS architectures

Таблица 3. Сравнение моделей классификации последовательности кадров

Table 3. Comparison of frame sequence classification models

Показатели сети Model markers	Рекуррентные нейронные сети Recurrent Neural Networks		Сеть на основе полносвязных слоев Network based on fully connected layers
	Сеть на основе GRU GRU-based network	Сеть на основе LSTM LSTM-based network	
Точность Accuracy, %	0,748	0,773	0,693
Время обработки, с Time of processing, s	0,19	0,18	0,06

По результатам, приведенным в табл. 3, делаем вывод, что наиболее эффективной моделью является модель, основанная на LSTM. Также стоит отметить, что достигнутое качество классификации не является конечным, так как требуется более тонкая настройка входных параметров нейронной сети. При ручном анализе данных было выявлено, что встречаются последовательности из трех кадров, где дымовое облако прослеживается только на одном из них, что, в свою очередь, негативно сказывается на конечном результате. Также необходимо отметить снижение ошибок идентификации пожароопасных объектов в видеопоследовательности в сравнение с подходом, анализирующим один кадр.

Сеть долгой краткосрочной памяти (LSTM)

Как было отмечено ранее, модель ИНС, основанная на LSTM, требует одновременной оптимизации параметров

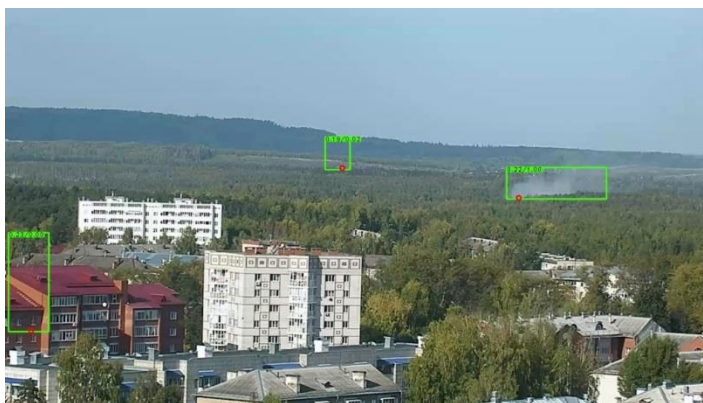
входного тензора и анализируемых кадров. Результаты оптимизации выходных параметров нейронной сети представлены в табл. 4.

Таблица 4. Настройка входных параметров сети на основе LSTM

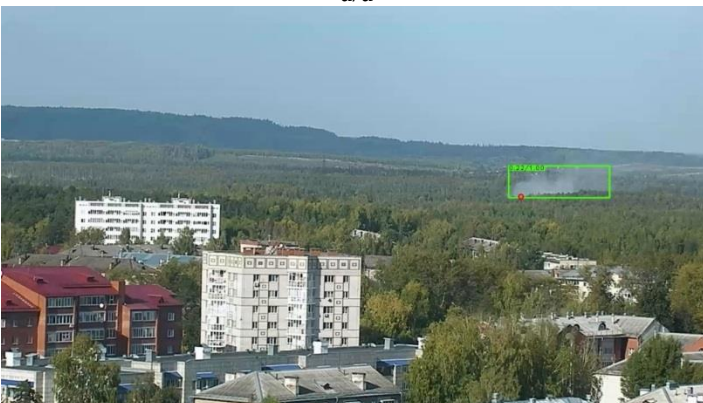
Table 4. Setting the input parameters of the LSTM-based network

Количество обрабатываемых кадров Number of processed frames	Размер кадра Frame size	Точность Accuracy, %	Время, с Time, s
3	28x28	68,98	0,04
	32x32	72,12	0,07
	64x64	75,05	0,11
	128x128	77,33	0,21
5	28x28	79,97	0,06
	32x32	83,21	0,08
	64x64	85,4	0,12
	128x128	85,7	0,25
7	28x28	73,74	0,18
	32x32	75,32	0,30
	64x64	78,5	0,41
	128x128	77,79	0,75
11	28x28	65,74	0,27
	32x32	68,54	0,49
	64x64	70,66	0,98
	128x128	71,03	1,12

В результате экспериментальной настройки входных параметров нейронной сети делаем вывод, что наибольшей эффективности модель достигает с размером входного тензора $5 \times 64 \times 64 \times 3$, конечная точность классификации видеопоследовательности составила 85,4 %. Результаты работы классификатора изображены на рис. 4.



а/а



б/б

Рис. 4. Визуализация работы классификатора: а) визуализация выделения объектов по модели обнаружения объектов; б) визуализация выделения объектов

Fig. 4. Visualization of the classifier work: а) visualization of object selection by the object detection model; б) visualization of object selection

Заключение

В данной статье были рассмотрены вопросы классификации изображений видеопоследовательности. В работе представлены результаты сравнения алгоритмов классификации одного кадра и их последовательности. Представлены результаты алгоритмов машинного обучения для классификации признаков дымового облака на одном кадре. Несмотря на высокие показатели точности – 0,836 %, представленные алгоритмы классификации допускают ошибки в идентификации пожароопасных объектов. Данные ошибки являются критичными, поэтому полученный результат не может считаться удовлетворительным. Указанные ошибки могут быть связаны с отсутствием в текущий момент времени в кадре дымового облака и отсутствием в анализе динамических признаков. Для решения данной проблемы рассмотрен подход, ос-

нованный на анализе видеопоследовательности. Представлены результаты сравнения различных моделей нейронных сетей для анализа видеопоследовательности. В результате анализа наибольшую эффективность показала модель на основе LSTM, точность которой составила 85,4 %. Также в статье приводится описание подбора входных параметров RCNN. На основе проведенного исследования подтверждена гипотеза о важности анализа динамически признаков дымового облака. В результате ошибка неверной классификации пожароопасных объектов минимизирована. Необходимо отметить, что предложенная модель нейронной сети не идеальна и допускает ошибки, связанные с неверной классификацией дождевых облаков, которые идентичны по текстурным признакам дымовым облакам. В целом работа системы оценивается экспертами как высокая.

СПИСОК ЛИТЕРАТУРЫ/REFERENCES

1. Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, Avidan S. Fast human detection using a cascade of histograms of oriented gradients. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). New York, NY, USA, 2006. pp. 1491–1498. DOI: 10.1109/CVPR.2006.119.
2. Matti Pietikäinen, Abdenour Hadid, Guoying Zhao, Timo Ahonen. Computer vision using local binary patterns. Berlin, Springer, 2011. 228 p.
3. Bay H., Tuytelaars T., Van Gool L. SURF: Speeded up robust features. Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science. Eds. A. Leonardis, H. Bischof, A. Pinz. Berlin, Heidelberg, Springer, 2006, vol. 3951, pp. 404–417. DOI: https://doi.org/10.1007/11744023_32
4. Laaksonen J., Oja E. Classification with learning k-nearest neighbors. IEEE International Conference on Neural Networks – Conference Proceedings. IEEE, 1996, vol. 3, pp. 1480–1483.
5. Zhao Jianhui, Zhang Zhong, Han Shizhong, Qu Chengzhang, Yuan Zhiyong, Zhang Dengyi. SVM based forest fire detection using static and dynamic features. Computer Science and Information Systems, 2011, vol. 8, Iss. 3, pp. 821–841. DOI: <https://doi.org/10.2298/CSIS101012030Z>.
6. Biau G., Scornet E. A random forest guided tour. Test, 2016, vol. 25, pp. 197–227. DOI: <https://doi.org/10.1007/s11749-016-0481-7>.
7. O'Shea K., Nash R. An introduction to convolutional neural networks. arXiv:1511.08458v2. 2015. DOI: <https://doi.org/10.48550/arXiv.1511.08458>
8. Rawat W., Wang Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural Computation, 2017, vol. 29, no. 9, pp. 2352–2449. DOI: 10.1162/NECO_a_00990
9. Recurrent Neural Networks: design and applications. Eds. L. Medsker, L.C. Jain. Boca Raton, CRC Press, 1999. 416 p. DOI: <https://doi.org/10.1201/9781003040620>.
10. Sundermeyer M., Schlüter R., Ney H. LSTM Neural Networks for language modeling. INTERSPEECH 2012. ISCA's 13th Annual Conference Portland, OR, USA, 2012. pp. 194–197. DOI: 10.21437/Interspeech.2012-65DOI:10.21437/Interspeech.2012-65
11. Laptev N.V., Laptev V.V., Gerget O.M., Kravchenko A.A., Kolpashchikov D.Yu. Visualization system for fire detection in the video sequences. Scientific Visualization, 2021, vol. 13, no. 2, pp. 1–9. DOI: 10.26583/sv.13.2.01.
12. Cloud AutoML Custom Machine Learning Models. Available at: <https://cloud.google.com/automl> (accessed: 18 May 2021).
13. Ramraj S., Nishant Uzir, Sunil R., Shatadeep Banerjee. Experimenting XGBoost algorithm for prediction and classification of different datasets. International Journal of Control Theory and Applications, 2016, vol. 9, no. 40, pp. 651–662.
14. Wang D., Zhang Y., Zhao Y. LightGBM: an effective miRNA classification method in breast cancer patients. ICCBB 2017: Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics. ICCBB 2017. New York, New York, USA, ACM Press, 2017. pp. 7–11. DOI: <https://doi.org/10.1145/3155077.3155079>
15. Liaw A., Wiener M. Classification and regression by RandomForest. R news, 2002, vol. 2/3, pp. 18–22.
16. Hancock J.T., Khoshgoftaar T.M. CatBoost for big data: an interdisciplinary review. Journal of Big Data, 2020, vol. 7, Article number: 94. DOI: 10.1186/s40537-020-00369-8
17. Shen B., Liu B.-D., Wang Q. Elastic net regularized dictionary learning for image classification. Multimed. Tools Appl., 2016, vol. 75, pp. 8861–8874.
18. Aakanksha Sharaff, Harshil Gupta. Extra-tree classifier with metaheuristics approach for email classification. Advances in Computer Communication and Computational Sciences, 2019, vol. 924, pp. 189–197. DOI: https://doi.org/10.1007/978-981-13-6861-5_17

Дата поступления: 20.02.2023 г.

Дата принятия: 01.06.2023 г.

Информация об авторах

Лаптев Н.В., инженер отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

Гергет О.М., доктор технических наук, ведущий научный сотрудник, Институт проблем управления им. В.А. Трапезникова Российской академии наук.

Лаптев В.В., инженер отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

Колпащиков Д.Ю., инженер отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

UDC 004.032.26

INVESTIGATION OF THE CAPABILITIES OF ARTIFICIAL NEURAL NETWORKS WHEN CLASSIFYING OBJECTS DYNAMIC FEATURES

Nikita V. Laptev¹,
nikitalaptev77@gmail.com

Olga M. Gerget²,
gerget@ipu.ru

Vladislav V. Laptev¹,
vvl39@tpu.ru

Dmitriy Yu. Kolpashchikov¹,
dyk1@tpu.ru

¹ National Research Tomsk Polytechnic University,
30, Lenin avenue, Tomsk, 634050, Russia.

² Institute of Control Sciences of Russian Academy of Sciences,
65, Profsoyuznaya street, Moscow, 117997, Russia.

Image classification is a classic machine learning task. Deep neural networks are widely used in the field of object classification. However, the problem of analyzing objects with dynamically changing features remains relevant. To solve this problem, the authors propose using a long short-term memory networks. Unlike classical convolutional neural networks, the proposed network uses information about the sequence of images, thereby providing a higher classification accuracy of detected objects with dynamic features. In the study, the authors analyze the classification accuracy of smoke cloud detection in a forest using various machine learning methods.

Key words: neural networks, traditional machine learning, classification, image, detection of fire hazards.

Information about the authors

Nikita V. Laptev, engineer, National Research Tomsk Polytechnic University.

Olga M. Gerget, Dr. Sc., leading researcher, Institute of Control Sciences of Russian Academy of Sciences.

Vladislav V. Laptev, engineer, National Research Tomsk Polytechnic University.

Dmitriy Yu. Kolpashchikov, engineer, National Research Tomsk Polytechnic University.

Received: 20 February 2023.

Reviewed: 1 June 2023.