Integrating Traditional Machine Learning and Neural Networks for Image Processing

Nikita Laptev [1], Vladislav Laptev [1], Olga Gerget [1] and Dmitriy Kolpashchikov [1]

[1] *Tomsk Polytechnic University, Street Lenina 30, Tomsk, 634050, Russia*

Abstract
The article describes a feasibility study to assess the use of neural networks and traditional machine learning algorithms to solve various problems including image processing. A brief description of some algorithms of traditional machine learning, as well as anautomated service for choosing the best method for a specific task, is given. The authors also describe the features of artificial neural networks and the most popular places for theirapplication. An algorithm for solving the problem of detecting fire hazardous objects andlocalizing a fire source in a forest using video sequence frames is presented. The article compares the characteristics of artificial neural network models according to the followingcriteria: underlying architecture, the number of analyzed frames, the size of the input image, the transfer learning model used as a feature vector composing network. Acomparative analysis of traditional machine learning algorithms and neural networks withlong short-term memory in the problem of classification of forest fire hazards is made. A solution to localization of the source of fire based on clustering is described. A hybrid algorithm for finding a fire source in a forest is developed and illustrated.

Keywords
Neural networks, traditional machine learning, classification, image, detection of fire hazards

1. Introduction

The most popular tasks in computer vision include: object detection, edge detection to segment regions of interest, classification and tracking. Such problems are successfully solved with machine learning algorithms. However, it should be noted that unlike neural networks, traditional algorithms, e.g. SVM [1,2], k-nearest neighbors [3,4], RandomForest [5,6], support vector machine, etc., operate under stringent conditions such as low variability of products, fixed position of the control object, high requirements for the quality of lighting. At the same time, the number of tasks that go beyond the capabilities of traditional machine learning is constantly growing. Such problems are commonly solved using neural networks, which proved efficient in image processing. Convolutional neural networks have significantly improved detection accuracy and increased the number of solvable tasks.

In this paper, the authors present a system for determining the source of fire in a forest. The problem of forest fires is relevant for the whole world. Advanced analysis of visual images obtained from fire detection systems, accurate localization and timely measures play a crucial role in preventing environmental disasters and minimizing environmental damage. In this regard, the importance of developing real time visual systems for accurate fire detection and forest fire location is beyond doubt.

Currently, there are many algorithms for detecting forest fires [7–9]. It should be noted that at present, a high quality of detection of research objects based on machine learning has been achieved, in particular, based on the "Object detection" technology [10, 11]. This article discusses a hybrid

approach to automated video monitoring of forest fires. To classify video stream data, the authors use a neural network model. A machine learning algorithm for localizing the source of fire.

Since the excitement for neural networks is often artificially induced, it is necessary to consider the feasibility of using neural networks or traditional machine learning algorithms when solving a certain type of problem. In this article, the authors propose to compare the results of the classification of these fire hazardous objects using models of artificial neural networks and machine learning algorithms. The choice of the optimal algorithm is carried out using the new cloud service AutoML [12], which allows automating the process of choosing machine learning algorithms.

2. Neural networks

Artificial intelligence systems have many directions of development. One of the most promising is neural networks. The neural network concept is to simulate a human's neural system. In particular, its ability to learn using previous experience and thus make fewer errors next time. This is the main feature of neural networks. The main tasks of neural networks are following:

- pattern recognition (classification) [13];
- regression [14];
- time series forecasting [15];
- segmentation [16];
- data generation [17].

The listed types of tasks cannot be considered strictly divided. One way or another they can flow into each other.

As mentioned earlier, the tasks solved by neural networks can be divided into types. The correct definition of task type is very important. Type of task defines neural network types suitable for it. The correct definition of task type is very important. Type of task defines neural network types suitable for it. Neural networks, in turn, can also be divided into types. Each type obtains a different outcome for some certain task.

Let us consider the first task from the list - classification. It is a very popular task whose goal define a class of the object. An object for classification could be anything. For example:

- image;
- video;
- audio;
- text;
- number series.

The first step of the workflow is preprocessing. The data should be prepared for a neural network. For example, we can transform data to the right shape or apply various filters to improve results.

Sometimes there are so many preprocessing algorithms that the task could be solved without neural networks. So why should use a neural network? The answer lies in the very essence of the neural network, namely in what "learning" is striving for - increasing the generalizing ability. Conclusion - neural networks are quite complex and require large resources, therefore, they should only be applied to large problems, where a simple algorithm or their combination will not be sufficient. Another disadvantage of using neural networks is resource intensity. The larger the image, the more informative features it can carry in itself, which means that the size of the neural network should be larger. The high computational cost is the reason why neural networks were not so popular even 10 years ago when images were processed by fully connected deep networks. Currently, convolutional neural networks are used for this type of task. The main idea of such networks is to train a small object called the convolution kernel, instead of matching each pixel - a neuron.

The next type of task is regression. Regression is estimating the relationships between a dependent variable and one or more independent variables. The regression task is not much different from the classification task. The basic idea is to predict the output vector based on the input vector. Often, regression problems without complex data types (e.g. video or image) can be solved using a simple perceptron.

For forecasting time series attention is focused on time. In this task, the temporal component is a key component. The perceptron or convolutional neural network is not able to easily solve it. But still,they can help. Let us take the video processing task as an example. The video sequence could be split into separate images. But for some tasks, a single frame does not carry a significant amount of information, so we should look into the whole sequence. There are several solutions for analyzing such data. The first is to consider each example separately, highlight the key features, combine them into a single object and work with it directly. The method is undoubtedly working, but not the best. The second is to use the Long Short-Term Memory (LSTM) network [18]. This type of network specially designedto analyze entire sequences of data. LSTM networks focus their attention on features that repeat not only from stack to stack of samples but also within a single stack of samples, which, directly, allows them to distinguish features of objects over time.

The segmentation task is mostly applicable to images and similar to the classification task. Many machine learning algorithms exist to solve this task: WaterShed [19], MeanShift [20]. However, traditional machine learning algorithms are not always enough. It is worth mentioning that neural networks solve this task with the same efficiency and sometimes even better [17,21].

Data generation task is also very popular nowadays. The large training set is very important for obtaining a high-quality prediction model. Neural networks could create new synthetic data based on available data. Popular solutions for this task: variational autoencoder [22], generative adversarial network [23].

3. Traditional Machine Learning

Traditional machine learning is a set of mathematical, statistical and computational methods for developing algorithms that can solve a problem not in a direct way, but based on finding patterns in a variety of input data. The solution is calculated not according to a formula, but according to the established dependence of the results on a specific set of features and their values.

Traditional machine learning differs from neural networks by feature extraction procedures. A neural network automatically extracts features in the process of training. The feature extracting capability defined by the model and architecture of the network. In traditional machine learning feature extraction carried out by feature detection algorithms, such as Scale Invariant Feature Transform (SIFT) [24], Speeded Up Robust Features (SURF) [25], Histogram of Oriented Gradients (HOG) [26] and etc. Feature extraction algorithm used to find key points and form a feature vector necessary for the further operation of the classification algorithms. These features can be corners, color schemes, image texture, etc.

SIFT

The SIFT algorithm uses a series of mathematical approximations to learn a representation of the image that is scale-invariant. In effect, it tries to standardize all images. This corresponds to the idea that if some feature (say a corner) can be detected in an image using some square-window of dimension σ across the pixels, then we would if the image was scaled to be larger, we would need a larger dimension kσ to capture the same corner. The mathematical ideas of SIFT are skipped, but the general idea is that SIFT standardizes the scale of the image then detects important key features. The existence of these features are subsequently encoded into a vector used to represent the image. Exactly what constitutes an important feature is beyond the scope of this paper.

SURF

The SURF algorithm is based on the same principles and steps as SIFT, but the details are different at each step. The algorithm has three main parts: points of interest detection, local area description, and matching. SURF works by finding a quick and rough approximation to the Gaussian difference using a technique called block blur. Block blur is the average value of all pixels in a given rectangle and can be calculated efficiently.

HOG

The main idea of the algorithm is that the appearance and shape of the object in the image can be described by the distribution of intensity gradients or the direction of the edges. The image is divided into small areas called cells. A histogram and directional gradient are compiled for the pixels in the cell data. To improve accuracy, local histograms are normalized for contrast by calculating a measure of

intensity over most of the image. The calculated value is applied to normalize all cells. This results in better invariance to changes in lighting and dimming.

After the procedures for the selection and extraction of features, a data classification model is built. There are many different classification algorithms such as SVM [27], k-nearest neighbors [28], RandomForest [5] and etc. The application of each algorithm is justified for a specific task, and the application requires a certain amount of experience. Therefore, the authors propose to use the methods of automatic machine learning (AutoMl) [12] to select the optimal algorithm.

At the moment, traditional computer vision algorithms more suitable for simple tasks when computing resources or dataset limited. Application of machine learning algorithm are following: robotics [29], augmented reality [30], automatic panorama stitching [31], virtual reality [32], 3D modeling [32], motion estimation [32], video stabilization [29], motion capture [32], video processing [29] and scene understanding [33].

4. Detection of Fire Hazardous Objects in Forest

Authors propose a combined algorithm to detect and localize fire hazardous objects in a forest. The algorithm uses a sequence of different steps:
1. Pre-processing algorithm subtraction frames for the extraction of dynamic features.
2. Object detection technology find fire hazardous objects on prepared data.
3. Post-processing algorithm.
4. Low-level descriptors allocate characteristic features of detected objects.
5. Machine learning algorithm classify detected objects in order to improve the detection accuracy.
6. Localization of the source of fire on the detected objects, using the clustering algorithm.

The use of preprocessing and post-processing algorithms, as well as the choice of neural network architecture, is described in detail in the article [34]. In the article, the authors note the high efficiency of detecting fire hazardous objects by neural networks. However, the system makes false-positive detection of objects with structures similar to smoke, such as fog and clouds. To solve this problem, the authors propose to introduce additional filtering of found objects by a classifier.

The detection system [34] converts the input image into grayscale and then subtracts the background. This causes the loss of some characteristic features necessary for the classification of the found objects. Due to this, the authors propose to use the original images for classification without using the frame pre-processing algorithm. Only part of the image detected by neural networks at stage 2 fed to the classifier.

To solve the problem of localization (point determination of the source) of the source of fire, the authors propose to use the classical method of traditional machine learning K-means [35]. This clustering algorithm successfully copes with the task and does not require a large amount of computing resources. It is also worth noting the ease of use of this algorithm.

5. Results

Two neural network architectures (see figure 1) and several traditional machine learning algorithms developed and tested. The first neural network model (see figure 1 "a") is a classifier based on transfer learning and the use of fully connected layers. During model training, we experimented with different model parameters: the size of the input tensor and the choice of a feature extraction model (see table 1). The second model (see figure 1 "b") is an improved version of the first model. Its key feature is the ability to analyze several frames at the same time, which is achieved using a combination of layers Time Distrubed and LSTM layers. In the process of model training, we experimented with model parameters: the number of frames fed to the network input (see table 2). Accuracy calculated as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP (True Positive) - smoke detected in the frame with smoke, FP (False Positive) - smoke detected in the frame without smoke, TN (True Negative) - smoke not detected in the frame without smoke, FN (False Negative) - smoke not detected in the frame with smoke. Accuracy calculated using

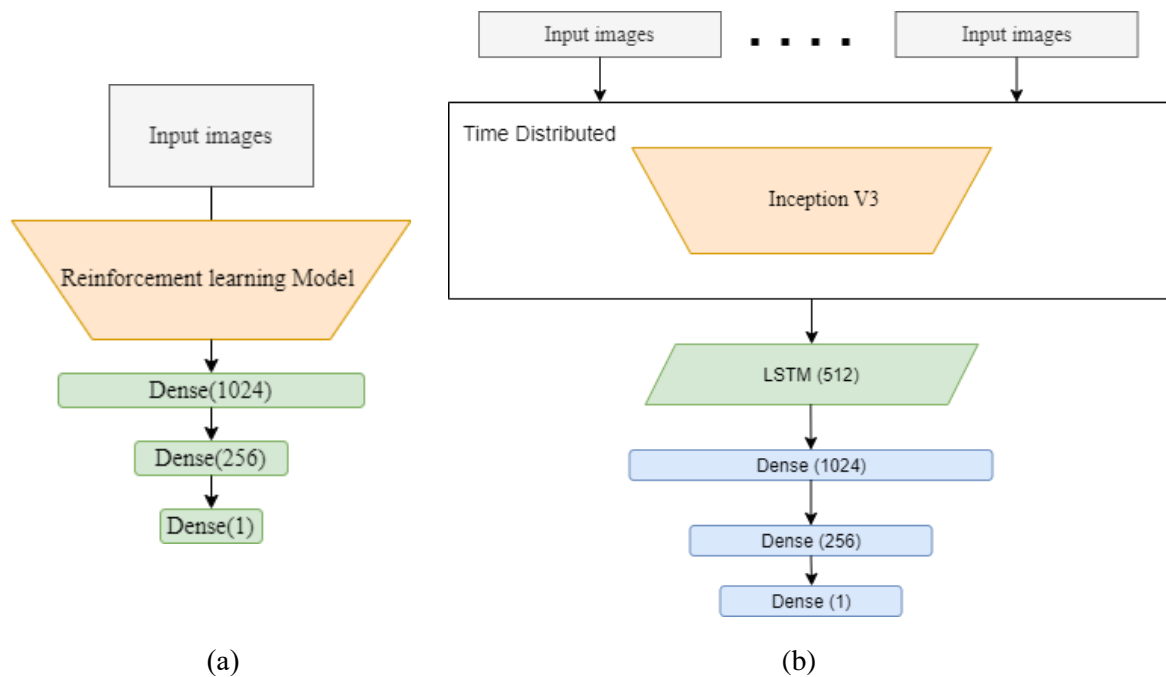390 images: 100 images contain smoke and 190 do not.

(a)   (b)

Figure 1: Neural network architectures

Table 1
Neural network architectures test result

Model name	Input size	Accuracy, %
MobileNetV2	28x28	43,1
	32x32	49,9
	64x64	58,3
	128x128	53,0
EfficientDet-D1	28x28	50,3
	32x32	57,8
	64x64	61,3
	128x128	60,0
ResNet50	28x28	57,4
	32x32	59,8
	64x64	60,6
	128x128	60,2
InceptionV3	28x28	53,2
	32x32	60,5
	64x64	64,7
	128x128	62,2

Table 2
Dependence of accuracy on the number of frames

Number of frames	Accuracy, %
3	75,0
5	85,7
7	78,5
11	71,3

The efficiency of traditional machine learning algorithms evaluated by EvalML library[36] of the AutoML web service. This library evaluates traditional machine learning algorithms using domain-specific objective functions. At the input, the algorithm receives a vector of features, which is formed as a map of the intensities of the image pixels, reduced to a strict format, or the result of the operation of one of the function descriptors. The test results are shown in Table 3.

Table 3
Results

Machine learning algorithms	Algorithm name	Accuracy, %
None	XGBoost	80,0
	LightGBM	76,8
	Random Forest	70,0
	CatBoost	44,2
	ElasticNet	45,8
	ExtraTree	66,5
HOG	XGBoost	61,8
	LightGBM	51,9
	Random Forest	57,1
	CatBoost	41,1
	ElasticNet	45,4
	ExtraTree	55,8

The best result with an accuracy 86% shown by a neural network based on a long short-term memory network (LSTM) using the InceptionV3 mode. The best result among traditional machine learning algorithms shown by the XGBoost algorithm without the use of additional function descriptors. The accuracy of XGBoost is 80%.

The accuracy metric MAE [37] is used to assess the quality of the K-means algorithm in the problem of localizing the fire source. As a result of testing, a metric score of 0.09 was obtained, which indicates that the predicted point is close to the annotated one. It should be noted that this algorithm works exclusively on the area being detected as a fire hazardous object. System operation is shown in Figure 2.
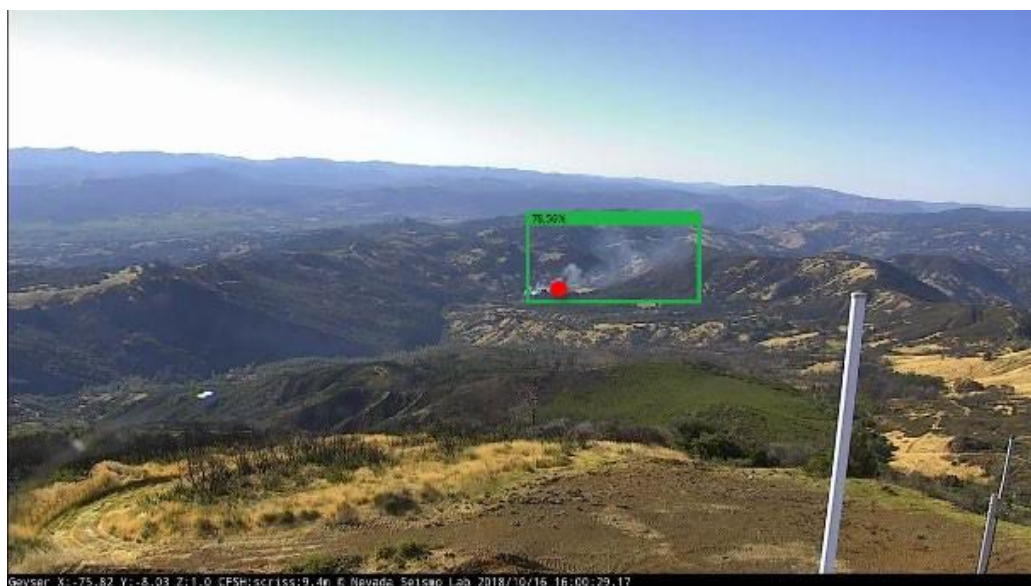

Figure 2: System performance visualization

6. Conclusion

This paper outlined the use of neural networks and traditional machine learning algorithms to solve automation problems. An algorithm to detect forest fire sources was further explained. The results of testing traditional machine learning algorithms and a neural network to classify data, namely, detected fire hazards, were presented. A long short-term memory (LSTM) neural network has shown the highest efficiency at the accuracy of 86%. It should be noted that the XGBoost traditional machine learning algorithm has also proved highly efficient with the accuracy of 80%. The results highlight the relevance of research underlying the selection of optimal algorithms for particular problems. Traditional machine learning algorithms can be used for automation problems with similar efficiency, while they require significantly less computational resources and do not require large databases as opposed to neural network models. These features were the main criteria for selection decisions on the localization of the source of fire. As a solution to this problem, it is proposed to use the clustering algorithm. However, neural networks ensure greater accuracy in image processing taking into account the dynamics based on raw data (data without preprocessing).

7. Acknowledgements

8. References

[1] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, T. Huang, Large-scale image classification: Fast feature extraction and SVM training, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2011: pp. 1689–1696. doi:10.1109/CVPR.2011.5995477.

[2] G.M. Foody, A. Mathur, The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM, Remote Sensing of Environment. 103 (2006) 179–189. doi:10.1016/j.rse.2006.04.001.

[3] P. Horton, K. Nakai, Better Prediction of Protein Cellular Localization Sites with the k Nearest Neighbors Classifier, 1997. www.aaai.org.

[4] O. Kramer, K-Nearest Neighbors, in: Springer, Berlin, Heidelberg, 2013: pp. 13–23. doi:10.1007/978-3-642-38652-7_2.

[5] A. Liaw, M. Wiener, Classification and Regression by RandomForest, R News. 2 (2002) 18–22.

[6] G. Biau, E. Scornet, A random forest guided tour, Test. 25 (2016) 197–227. doi:10.1007/s11749-016-0481-7.

[7] J. Zhao, Z. Zhang, S. Han, C. Qu, Z. Yuan, D. Zhang, SVM based forest fire detection using static and dynamic features, Computer Science and Information Systems. 8 (2011) 821–841. doi:10.2298/CSIS101012030Z.

[8] V. Vipin, Image processing based forest fire detection, International Journal of Emerging Technology and Advanced Engineering 2.2. 8 (2012) 87–95. www.ijetae.com.

[9] C. Emmy Prema, S.S. Vinsley, S. Suresh, Multi Feature Analysis of Smoke in YUV Color Space for Early Forest Fire Detection, Fire Technology 2016 52:5. 52 (2016) 1319–1342. doi:10.1007/S10694-016-0580-8.

[10] D. Alexandrov, E. Pertseva, I. Berman, I. Pantiukhin, A. Kapitonov, Analysis of machine learning methods for wildfire security monitoring with an unmanned aerial vehicles, *Conference of Open Innovation Association, FRUCT*. 2019-April (2019) 3–9. doi:10.23919/FRUCT.2019.8711917.

[11] V. Danilov, O. Gerget, K. Klyshnikov, E. Ovcharenko, A. Frangi, Comparative study of deep learning models for automatic coronary stenosis detection in x-ray angiography, CEUR Workshop Proceedings. 2744 (2020). doi:10.51130/GRAPHICON-2020-2-3-75.

[12] Cloud AutoML Custom Machine Learning Models | URL: https://cloud.google.com/automl.

[13] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet Classification with Deep Convolutional Neural

Networks, Advances in Neural Information Processing Systems. (2012) 1097–1105.

[14] D.F. Specht, A general regression neural network, IEEE Transactions on Neural Networks. (1991) 568–576.

[15] B. Zoph, Q. V. Le, Neural Architecture Search with Reinforcement Learning, 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings. (2016).

[16] V. V. Danilov, I.P. Skirnevskiy, R.A. Manakov, O.M. Gerget, F. Melgani, Feature selection algorithm based on PDF/PMF area difference, Biomedical Signal Processing and Control. 57 (2020). doi:10.1016/j.bspc.2019.101681.

[17] V. V Danilov, O.M. Gerget, D.Y. Kolpashchikov, N. V Laptev, R.A. Manakov, L.A. Hérnandez-Gómez, F. Alvarez, M.J. Ledesma-Carbayo, BOOSTING SEGMENTATION ACCURACY OF THE DEEP LEARNING MODELS BASED ON THE SYNTHETIC DATA GENERATION, Noa.Gwlb.De. (2021). doi:10.5194/isprs-archives-XLIV-2-W1-2021-33-2021.

[18] F. Karim, S. Majumdar, H. Darabi, S. Chen, LSTM Fully Convolutional Networks for Time Series Classification, IEEE Access. 6 (2017) 1662–1669. doi:10.1109/ACCESS.2017.2779939.

[19] H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh, W.L. Nowinski, Medical image segmentation using k-means clustering and improved watershed algorithm, in: Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation, 2006: pp. 61–65. doi:10.1109/ssiai.2006.1633722.

[20] W. Tao, H. Jin, Y. Zhang, Color image segmentation based on mean shift and normalized cuts, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 37 (2007) 1382–1389. doi:10.1109/TSMCB.2007.902249.

[21] P. Moeskops, M.A. Viergever, A.M. Mendrik, L.S. De Vries, M.J. Benders, I. Išgum, Automatic segmentation of MR brain images with a convolutional neural network, IEEE Transactions on Medical Imaging. 35 (2016) 1252–1261.

[22] V. Laptev, V. Danilov, RESEARCH OF VARIATION AUTOENCODER FOR SYNTHESIS OF NEW MEDICAL DATA, Collection of Selected Articles of the Scientific Session of TUSUR. (2020) 68–70.

[23] V. Laptev, V. Danilov, O. Gerget, RESEARCH OF GENERATIVE-COMPOSITIVE NETWORKS FOR SYNTHESIS OF NEW MEDICAL DATA, Automation and Simulation in Design and Control. (2020).

[24] T. Lindeberg, Scale Invariant Feature Transform, Scholarpedia. 7 (2012) 10491. doi:10.4249/scholarpedia.10491.

[25] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded up robust features, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, Berlin, Heidelberg, 2006: pp. 404–417. doi:10.1007/11744023_32.

[26] A. Aggarwal, S. Biswas, S. Singh, S. Sural, A.K. Majumdar, Object tracking using background subtraction and motion estimation in MPEG videos, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006: pp. 121–130. doi:10.1007/11612704_13.

[27] Y. Chen, X.S. Zhou, T.S. Huang, One-class SVM for learning in image retrieval, in: IEEE International Conference on Image Processing, IEEE, 2001: pp. 34–37. doi:10.1109/icip.2001.958946.

[28] J. Laaksonen, E. Oja, Classification with learning k-nearest neighbors, in: IEEE International Conference on Neural Networks - Conference Proceedings, IEEE, 1996: pp. 1480–1483. doi:10.1109/icnn.1996.549118.

[29] X. Zhang, J.Y. Lee, K. Sunkavalli, Z. Wang, Photometric Stabilization for Fast-forward Videos, Computer Graphics Forum. 36 (2017) 105–113. doi:10.1111/cgf.13276.

[30] H. Abu Alhaija, S.K. Mustikovela, L. Mescheder, A. Geiger, C. Rother, Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes, International Journal of Computer Vision. 126 (2018) 961–972. doi:10.1007/s11263-018-1070-x.

[31] G. Meneghetti, M. Danelljan, M. Felsberg, K. Nordberg, Image alignment for panorama stitching in sparsely structured environments, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2015: pp. 428–439. doi:10.1007/978-3-319-19665-7_36.

[32] T. Alldieck, M. Kassubeck, B. Wandt, B. Rosenhahn, M. Magnor, Optical flow-based 3D human

motion estimation from monocular video, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2017: pp. 347–360. doi:10.1007/978-3-319-66709-6_28.

[33] B. Zheng, Y. Zhao, J. Yu, K. Ikeuchi, S.-C. Zhu, C. by Derek Hoiem, J. Hays, J. Xiao, A.B. Khosla Zheng, K. Ikeuchi, Y. Zhao, J. Yu, S. Zhu, Scene Understanding by Reasoning Stability and Safety, Int J Comput Vis. 112 (2015) 221–238. doi:10.1007/s11263-014-0795-4.

[34] N. Laptev, V. Laptev, O. Gerget, A. Kravchenko, Visualization System for Fire Detection in the Video Sequences, Sv-Journal.Org. 13 (2021) 1–9. doi:10.26583/sv.13.2.01.

[35] A. Likas, N. Vlassis, J. J. Verbeek, The global k-means clustering algorithm, Pattern Recognition. 36 (2003) 451–461. doi:10.1016/S0031-3203(02)00060-2.

[36] EvalML is an AutoML library written in python., URL: https://github.com/alteryx/evalml.

[37] C.J. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Climate Research. 30 (2005) 79–82. doi:10.3354/CR030079.