Article

Deep Learning Method for Recognition and Classification of Images from Video Recorders in Difficult Weather Conditions

Aleksey Osipov ¹ , Ekaterina Pleshakova ², Sergey Gataullin ² , Sergey Korchagin ² , Mikhail Ivanov ² , Anton Finogeev ^{3,*} and Vibhash Yadav ⁴

¹ Information Security Department, Financial University under the Government of the Russian Federation, 4-th Veshnyakovsky Passage, 4, 109456 Moscow, Russia; avosipov@fa.ru

² Department of Data Analysis and Machine Learning, Financial University under the Government of the Russian Federation, 4-th Veshnyakovsky Passage, 4, 109456 Moscow, Russia; espleshakova@fa.ru (E.P.); stgataullin@fa.ru (S.G.); sakorchagin@fa.ru (S.K.); mnivanov@fa.ru (M.I.)

³ CAD Department, Penza State University, 440026 Penza, Russia

⁴ Department of Information Technology, Rajkiya Engineering College, Atarra, Banda 210201, India; vibhashds10@recbanda.ac.in

* Correspondence: fanton3@yandex.ru; Tel.: +7-89270902350

Citation: Osipov, A.; Pleshakova, E.; Gataullin, S.; Korchagin, S.; Ivanov, M.; Finogeev, A.; Yadav, V. Deep Learning Method for Recognition and Classification of Images from Video Recorders in Difficult Weather Conditions. *Sustainability* **2022**, *14*, 2420. <https://doi.org/10.3390/su14042420>

Academic Editors: Francesco Domenico Moccia and Alessandro Sgobbo

Received: 8 January 2022

Accepted: 14 February 2022

Published: 20 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The sustainable functioning of the transport system requires solving the problems of identifying and classifying road users in order to predict the likelihood of accidents and prevent abnormal or emergency situations. The emergence of unmanned vehicles on urban highways significantly increases the risks of such events. To improve road safety, intelligent transport systems, embedded computer vision systems, video surveillance systems, and photo radar systems are used. The main problem is the recognition and classification of objects and critical events in difficult weather conditions. For example, water drops, snow, dust, and dirt on camera lenses make images less accurate in object identification, license plate recognition, vehicle trajectory detection, etc. Part of the image is overlapped, distorted, or blurred. The article proposes a way to improve the accuracy of object identification by using the Canny operator to exclude the damaged areas of the image from consideration by capturing the clear parts of objects and ignoring the blurry ones. Only those parts of the image where this operator has detected the boundaries of the objects are subjected to further processing. To classify images by the remaining whole parts, we propose using a combined approach that includes the histogram-oriented gradient (HOG) method, a bag-of-visual-words (BoVW), and a back propagation neural network (BPNN). For the binary classification of the images of the damaged objects, this method showed a significant advantage over the classical method of convolutional neural networks (CNNs) (79 and 65% accuracies, respectively). The article also presents the results of a multiclass classification of the recognition objects on the basis of the damaged images, with an accuracy spread of 71 to 86%.

Keywords: Canny edge detector; deep learning; difficult weather conditions; image classification; histogram-oriented gradient method; bag-of-visual-words; back propagation neural network

1. Introduction

The trend in the development of modern urbanism is the transition to digital technologies to improve the efficiency of managing complex distributed objects in urban environments. The main goal is to achieve the sustainable functioning of the systems for ensuring the life of the urban population. Big data technologies, data mining, deep learning, and predictive analytics are of particular importance for the implementation of the Smart Sustainable City concept. The transition to the concept of the sustainability of the urban environment requires the development of proactive intelligent systems that are designed to prevent the risks of the occurrence and development of critical events at the distributed infrastructure facilities of the urban environment, which include the

engineering and technical networks (electrical, thermal, gas distribution, water and sewer networks, oil pipelines, etc.) and the urban transport system.

As an example of a distributed system in which many critical events occur every minute, consider an urban transport network. The objects of monitoring are the road sections, the elements of the road infrastructure, the vehicles, and the pedestrians. The sustainable functioning of this system requires solving the problems of identifying and classifying road users in order to predict the likelihood of traffic accidents and prevent emergency situations [1,2].

More recently, unmanned vehicles have appeared on our roads, as long as they are operated under human control. This is understandable both from the technical and legal points of view. The introduction of unmanned vehicles will unload roads, improve the environment, and optimize urban spaces [3–8]. However, the emergence of unmanned vehicles on city highways significantly increases the risks of critical events. To minimize them, and to improve road safety, intelligent transport systems, embedded computer vision systems, video surveillance systems, and photoradar systems are used. Therefore, the primary task is to collect and analyze the data from the various sources, which is obtained primarily from many similar devices.

There are several tasks that need to be solved before the total introduction of drones. The issue of safety is key in self-driving cars. Whether self-driving cars will be allowed on the roads depends on an effective solution [9–11]. The solution to this problem is multi-criteria and comes down to achieving several tasks, one of which is the timely detection and identification of objects and obstacles. The main problem here is the recognition and classification of objects and critical events in difficult weather conditions. For example, water droplets, snow, dust, and dirt on camera lenses make images less accurate in object identification, license plate recognition, vehicle trajectory detection, etc. The key parameter in this is the speed of an autonomous vehicle. The higher the speed, the faster the evaluation of the image must be made so that the vehicle control system can react and take the necessary actions.

To solve these problems, methods for sequential image processing are proposed, with a consideration to the analysis of their effectiveness when working in bad weather conditions. We obtained the best results, which proves the importance and novelty of the results of the research. A multiclass classification method is proposed, which is the next new and important result.

The article consists of six sections. In the Section 1, “Introduction”, the problem and purpose of the study are formulated. The Section 2 provides a theoretical overview of the research on this topic, identifies unsolved problems, and sets up the problem of object recognition in images from video cameras obtained in difficult weather conditions using the selected promising methods. The Section 3, “Materials and Methods”, provides a brief description of the proposed methods for solving the problems of the recognition and classification of noisy images. The Section 4, “Experiments and Results”, analyzes the method of image preprocessing, provides the indicators for the data preparation and evaluation, and presents the results. The Section 5, “Discussion”, presents the interpretation of the results. In the Section 6, “Conclusion”, the results are summarized, and the direction of future research is determined.

2. Theoretical Background

The most effective approaches for detecting obstacles in individual images are the approaches that use deep neural networks [12–15].

Especially popular among these at the moment are CNNs, which use a type of machine learning in which the model learns to perform classification tasks directly on the image [16,17]. Using CNNs has several advantages: thanks to deep neural networks, we have a flexible architecture with the addition of layers of neurons, which increases the ability of these networks to learn [18]; CNNs can be retrained to perform new recognition tasks, which allows existing networks to be used; and CNNs have high accuracy and reliability.

Recently, the performances of CNNs have been significantly improved [19–25]. When using this type of neural network in combination with powerful graphics-processing units [26], the CNN is the key technology behind new developments in driverless driving and facial recognition. However, as the authors of [27] note, convolutional neural networks work very slowly with high-resolution images and on devices with weak processors. This is explained by the fact that, in order to obtain an acceptable field susceptibility with convolutional layers, it is necessary to use large kernels (for example, 7×7 or 9×9), or a large number of layers [28], and this requires large computational resources. To avoid this, most existing systems are limited to image sizes smaller than 41×41 (pixels).

In addition, most modern CNN-based feature classifiers use downsampled convolutional feature maps, which lose much of the spatial information, and thus do a poor job of classifying small features. The serious obstacles for convolutional networks include blurring, flare, and a significant change in scale in uncontrolled conditions.

Another modification of deep neural networks is the YOLO (you only look once) line of single-pass detectors. The authors of [29] applied this type of detector to solve the problem of registering vehicles passing through toll portals in real time. The authors found that the Tiny YOLOv3 showed 100.0% recall and 98.5% accuracy.

Meghana Dinesh Kumar et al. [30] used another classifier type that is based on the BoVW model, which is actively used in image classification. Moreover, studies have shown that the “bag-of-visual-words” schemes for classifying histopathological images showed a high accuracy, which was 96.50%.

The speed of the image processing and the quality of the object classification are greatly improved when images are subjected to preprocessing steps. The image-to-image conversion is implemented using image filters [31–34], thresholding operations [29,30,35–37], morphological operations [31,38], and artificial neural networks [39,40].

The Canny method is actively used in computer vision to determine the boundaries of an image. It is less sensitive to noise, and by using a Gaussian filter, it attenuates them well. Canny optimization algorithm-based technology is widely used because of its good signal-to-noise ratio and detection accuracy. The traditional Canny operator is not without flaws: it does not have the adaptive ability to choose the variance of the Gaussian filter. Filtering requires human intervention, and the choice of the Gaussian filtering parameter affects the edge retention and the noise reduction effect [41]. On the basis of the traditional Canny operator, there are many improved Canny operators [42]. For example, in [43], the authors propose a method for determining the germination of potatoes on the basis of a multispectral image in combination with a supervised multithreshold segmentation model (SMTSM) and a Canny edge detector. The authors of [44] effectively eliminate noise using the method of combining global and local edge detection to extract the edge. By using the proven algorithm [45], which consists of five steps (the weighted mean quaternion filter; the Sobel vector gradient calculation; the interpolation-based nonmaximum suppression; the edge detection; and the connection), you can also work with color images. Low image contrast can be dealt with by using an improved method that is based on the Canny algorithm, which is as follows: two adaptive thresholds are obtained by performing a differential operation on the amplitude gradient histogram; then, the edge points are connected to obtain some generalized chains, after which it is necessary to calculate the mean value to remove the generalized chains that are less than the mean value; and finally the image edge detection results are obtained by linear fitting [46].

Der-Chang Tseng et al. [31] propose a hybrid approach to reconstructing an image damaged by random noise. The proposed method uses morphological component analysis (MCA) to decompose an image into the texture, structure, and the edge parts. Then, the block-matching and 3D-filtering (BM3D) method, the ANLM scheme, and the K-SVD algorithm are used to eliminate the noise in the texture, structure, and edge portions of the image, respectively. The results of the experiments show that the proposed approach makes it possible to effectively eliminate random noise in various parts, while, at the same time, a degraded image can be well restored.

The classical edge detection methods of Robert, Sobel, and Pruitt work with the pixels of the neighboring areas and obtain a gradient with pattern approximation, which is relatively simple and easy to implement and has good real-time performance; however, these operators are sensitive to noise, and their edge accuracy needs to be improved. The authors of [37] note that using the Sobel filter to detect COVID-19 using X-ray images improved the performance of a convolutional neural network. They rated this set of methods as the best of the wide range of options studied.

Detection becomes more difficult under adverse weather conditions, and the authors, Rachel Blin et al. [47], use polarization-encoded images in combination with classical training methods, namely, the DPM (deformable part model) and HOG methods. The authors note that polarimetry, combined with deep learning, can improve the performance by about 20–50% when performing various detection tasks.

On the basis of the material presented above, we have selected the most promising methods for solving our problem. These include the Canny image preprocessing method, and the BoVW-based classification method. For comparison, the CNN method was used. By using these methods, we were able to recognize objects from the images taken by video cameras through a protective glass covered with drops of water and dirt. When using other methods, including convolutional neural networks, such images are rejected.

3. Materials and Methods

3.1. Finding an Edge in a Nonblurred Image Area

Our first task was to find an algorithm to detect the required areas. We decided to use a Canny edge detector. The Canny edge detector uses a Gaussian filter to smooth the image and remove noise [48–52]. For the Gaussian filter kernel, the size equation is applied $((2k + 1) \times (2k + 1))$ and is given by the Canny edge detector:

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i - (k + 1))^2 + (j - (k + 1))^2}{2\sigma^2}\right); 1 \leq i, j \leq (2k + 1) \quad (1)$$

where H_{ij} is the value of the matrix element with coordinates, i, j (upper left matrix element, H_{11}); σ is the Gaussian standard deviation (in pixels); and k is the aperture radius, i.e., the size of the square aperture $((2k + 1) \times (2k + 1))$, respectively.

The image intensity gradients were found with the Canny edge detector, and the image intensity gradients were found by applying nonmaximum suppression in order to dispose of the edge detection spurious response. The Canny edge detector then uses a double threshold to determine the potential edges of the edge detector [53–58]. We used 80 for the first threshold, and 200 for the second. The edge detection then ends and suppresses all the other edges that are weak and that are not related to the strong edges.

We looked at our CSV file for the training dataset (Figure 1). By using both the picture and CSV files, we were able to highlight the required area on the photo.

However, we had “bad” photos, on which it was impossible to detect the required area, even for a human being. The main idea is that it is possible to find edges only in the area where the image is not blurred. This algorithm worked quite well on our image sample (Figure 2).

Since the black color on the picture is zero, and since we were working with a gray image with only one channel, we were able to find the first non-zero elements from the left, the right, and up and down, and to detect the required area. However, 4.52% of our photos had full black edge images, and so we did not train on them.

3.2. Applying the IoU Metric to Transform Our Images with True and Predicted Areas

Then, we needed to somehow score our algorithm. We decided to use the intersection-over-union metric [59–63] (Figure 3).

	filename	center_x	center_y	size	label
0	571.jpeg	1746	691	138	1
1	387.jpeg	1132	649	100	0
2	813.jpeg	166	642	108	0
3	1351.jpeg	1693	753	104	1
4	1367.jpeg	1320	647	112	1
...	...	...	...	...	...
2451	379.jpeg	634	697	130	0
2452	648.jpeg	508	633	102	1
2453	1700.jpeg	534	636	64	0
2454	691.jpeg	937	685	150	0
2455	1677.jpeg	1885	749	68	1

Figure 1. Snippet dataset.

Figure 2. Canny edge detector operation: (1) image sample; (2) edges on the image; (3) edge image; and (4) transformed image.

Figure 3. IoU metric.

The IoU is defined as $\frac{\text{Area of Overlap}}{\text{Area of Union}}$. To perform this metric, we transformed our images with the true and predicted areas into the following format: the required area was drawn in a white color, and another part of the image was drawn in black [64–66].

3.3. Using a Descriptor

The HOG method is based on the assumption that the type of distribution of the image intensity gradients makes it possible to accurately determine the presence and shape of the objects present on it.

The image is divided into cells. Histograms of the h_i -directed gradients of the internal points are calculated in the cells. They are combined into one histogram ($h = f(h_1, \dots, h_k)$), after which the image is normalized in brightness. The normalization multipliers can be obtained in several ways, but they show approximately the same results. We used the following:

$$h_L = \frac{h}{\sqrt{\|h\|_2^2 + \zeta^2}} \quad (2)$$

where $\|h\|$ is the norm used, and ζ is some small constant.

When calculating the intensity gradients, the image is convolved with the kernels, $[-1, 0, 1]$ and $[-1, 0, 1]^T$, and, as a result, matrices of the derivatives, D_x and D_y , will be formed along the x and y axes. The matrices, D_x and D_y , are used to calculate the angles and values at each point of the image.

Figure 4 shows the result of applying the HOG method to images of machines obtained through a clear and water-covered protective glass. A drop of water leads to a blurring of the image (Figure 4(2)) and a decrease in the gradient in the corresponding HOG region. In Figure 4(4), these are the dark areas inside the histogram.

3.4. Image Classification Using BoVW

To improve the performances of the descriptors, it is advisable to use the BoVW method [67]. This approach considers the blocks as key parts of the plant, and each block's HOG represents the local information of the corresponding part. Next, we cluster the HOGs of all the blocks in the training set into homogeneous groups using K-means, and the centers will be the mean values of the blocks' HOGs within the cluster.

Figure 4. The result of applying the HOG method to images of machines: (1) the image of the machine obtained through a clear protective glass; (2) the image of the machine obtained through glass covered with water droplets; (3) the image of Machine 1 processed by the HOG method; and (4) the image of Machine 2 processed by the HOG method.

3.5. CNN Image Classification

The CNN is a multistage feed-forward learning architecture, where each stage contains multiple layers that are used to tune the network [68]. For classification problems, the CNN accepts the input data in the form of images. These images go through a series of convolutional and subsampling layers to extract features. Fully connected layers classify an image according to certain outcomes, such as the accuracy and loss [69].

In this article, we used a special case of a CNN with the included VGG-16 structure [70]. It consists of five different blocks that are installed in series so that the output of each block is defined as the input of the next block (see Figure 5). In this architecture, the network extracts properties, such as the textures, shapes, and colors, from the input images.

Figure 5. Network architecture including VGG-16 structure and modified classifier block.

The VGG-16 contains 13 convolutional layers, with 3×3 kernel sizes, and five maximal pooling layers, with a 2×2 window. The activation function for each convolutional layer is the ReLU (Rectified Linear Unit) and it performs the following mathematical operation on each input:

$$f(x) = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

To diagnose the various types of objects encountered on the road, a modified model was used that classifies these objects. The modified model is followed by a block classifier. This block contained two maximum pooling layers, with a 2×2 window. After each max-pooling layer, a dropout layer is placed for regularization. Then comes the flatten layer (for smoothing the output), batch normalization, and the fully connected world. A fully connected world contains several neurons, depending on the number of classes used.

4. Experiments and Results

4.1. Image Preprocessing

An important step in determining objects on the road is the preprocessing of the images from video recorders. In this work, two types of images are processed: the first type of image is taken through clean protective glass; and the second is taken through glass covered with drops of water or dirt.

Figure 6 shows the conversion of a video camera image into a converted image. The camcorder shoots through a protective glass covered with small drops of water or dirt, which accordingly leads to a partial blurring of the boundaries of the objects. As can be seen in Figure 6(1), the Canny detector will ignore the blurry boundary. As can be seen in Figure 6(3), in the transformed image, the corresponding block will be removed from further consideration (Figure 6(4)).

Figure 6. Processing images from a video camera: (1) raw image; (2) grayscale image; (3) Canny processed image; and (4) transformed image.

In addition to cars, trees next to the road, billboards, etc., fell under the white blocks of the converted image. Particular attention should be paid to road signs, road markings, and roadway boundaries.

4.2. Data Preparation and Evaluation Metrics

To train the models, we used 1700 photos taken with a video recorder through a clean protective glass, and 830 photos taken through a protective glass covered with drops of water or dirt. Both sets of photographs were cropped with images of cars, people, and other objects on the road. The numbers of the selected samples are shown in Table 1.

Table 1. Numbers of samples used for training and testing computer vision methods for recognition and classification of objects on the road.

Objects	Images Were Taken through Clear Protective Glass		Images Were Taken through Protective Glass Covered with Drops of Water or Dirt	
	Multiclass Classification	Binary Classification	Multiclass Classification	Binary Classification
Cars	1942	1942	360	360
Pedestrians	1856	1856	345	345
Road signs	1125		385	
Traffic lights	830		320	
Pedestrian crossings	985		240	
Road markings	1430		380	

Each image was scaled to 70×70 and was classified.

For images taken through a clear protective glass, the samples were divided into “training” and “testing” in a ratio of 4 to 1. Figure 7 shows an example of the binary classification: 0 is a machine, and 1 is a person.

Figure 7. Images and labels.

The resulting numbers of the object images taken through protective glass covered with drops of water or dirt is significantly less than in the case of the images taken through clear protective glass, and so a five-fold cross-checking method was applied to them. The data was divided into two parts. The training sample relates to the testing sample as 80 to 20% of all of the data, respectively. The distribution structures of the training and testing data are shown in Figure 8.

Figure 8. Illustration of the five-fold method used to evaluate the effectiveness of the proposed methods.

For each fold, 80% of the data was used for training, while the rest were evaluated during the testing phase. The data used in the test phase were shifted in each fold.

4.3. Intelligent System

To develop an intelligent system, we tested two methods: a CNN and a HOG–BoVW–BPNN. The comparisons were made on the basis of the two models. The first model uses a binary classification: a machine is a person, and the second is multiclass.

We used the CNN structure presented in Section 3.5, with two neurons in a fully connected layer.

We trained on the Tesla K80. We used the cross-entropy loss as the loss function, and an Adam optimizer with a learning rate of 1×10^{-4} . The model was trained for 100 epochs.

The results of the classification accuracy of the proposed CNN model are presented in the normalized confusion matrices in Figures 9 and 10.

Figure 9. Confusion matrix for two samples processed by the CNN method. The first sample was taken through clean protective glass, and the second one was taken through a glass covered with small drops of water or dirt.

		First selection		Second selection	
True label	cars	0.88	0.12	0.80	0.20
	pedestrians	0.16	0.84	0.23	0.77
		cars	pedestrians	cars	pedestrians
		Predicted label			

Figure 10. Confusion matrix for two samples processed by the HOG–BoVW–BPNN method. The first sample was taken through clean protective glass, and the second one was taken through a glass covered with small drops of water or dirt.

From this matrix, one can observe the true-positive rates (the diagonal values) and the false-positive rates (the entries in columns other than the diagonals) of each class.

When implementing the HOG–BoVW–BPNN method, a visual dictionary was formed, which consisted of 500 visual words. The BoVW vocabulary was created from the samples that were used in the CNN training.

The results of the classification accuracy of the proposed HOG–BoVW–BPNN model are presented in the normalized uncertainty matrix in Figure 11, and the CNN models are presented in Figure 12.

True label	cars	0.78	0.18	0.00	0.00	0.02	0.02
	pedestrians	0.20	0.77	0.03	0.00	0.00	0.00
	road signs	0.00	0.02	0.84	0.11	0.00	0.03
	traffic lights	0.00	0.01	0.13	0.86	0.00	0.00
	pedestrian crossings	0.02	0.00	0.00	0.00	0.71	0.27
	road markings	0.02	0.01	0.00	0.00	0.16	0.81
		cars	pedestrians	road signs	traffic lights	pedestrian crossings	road markings
		Predicted label					

Figure 11. Confusion matrix for the proposed HOG–BoVW–BPNN classification model.

True label	cars	0.62	0.24	0.07	0	0.03	0.04
	pedestrians	0.22	0.64	0.08	0.04	0	0.02
	road signs	0	0.04	0.76	0.16	0	0.04
	traffic lights	0	0.04	0.16	0.8	0	0
	pedestrian crossings	0.03	0.01	0.01	0	0.67	0.28
	road markings	0.02	0.02	0.01	0	0.19	0.76
		cars	pedestrians	road signs	traffic lights	pedestrian crossings	road markings
		Predicted label					

Figure 12. Confusion matrix for the proposed CNN classification model.

5. Discussion

Over the past ten years, deep machine learning methods have made significant progress in the field of object classification. Convolutional neural networks, decision trees, etc., perform this task much faster and more accurately than a human. However, we have drawn attention to some of the limitations of these methods that are associated with their applications in real conditions. For example, the task of classifying vehicles moving in a stream and blocked by other vehicles is poorly solved using the convolutional neural networks that are commonly used today. As another example, note that most computer vision methods trained on objects photographed under ideal conditions lose classification accuracy when the processing images are distorted by water droplets or dirt. Yun Wei et al. [71] used the Harr and HOG methods to improve the quality and speed of the image recognition and classification.

The use of convolutional neural networks for processing such images leads to a large loss of accuracy. Therefore, in order to improve the accuracy, a variant of the BoVW implementation was chosen, which has proven itself to be adept at solving the computer vision problems in the HOG–BoVW–BPNN. In order to improve the performance of the method, the images were preprocessed using the Canny operator, which makes it possible to ignore their noisy and damaged parts. In the rest of the image, the objects were identified and classified using the HOG–BoVW–BPNN algorithm. The principle of the operation of this algorithm makes it possible to classify an object, even if part of its image is lost. Minor damage to the object image has little effect on the classification result.

When comparing the two binary classifiers, the HOG–BoVW–BPNN and the CNN, while working on the damaged images of cars and pedestrians, we obtained a significant advantage for the latter over the former (accuracies of 79 and 65%, respectively), which proves the importance and novelty of the results of the studies performed.

Moreover, we compared our results in the case of the clear safety glass with those from J. Chitra et al. [72]. In this work, two methods, a CNN and a HOG-support vector machine (SVM), were compared. It turns out that the results for the CNN coincided to within 1%. When comparing the HOG–BoVW–BPNN and HOG–SVM methods, the former proved to be approximately 5% more accurate. Unfortunately, we did not find similar studies with images taken through a protective glass covered with water droplets or dirt in the literature.

we studied. However, it is not only pedestrians and cars that fall into the field of action of the video camera. Of particular interest are the objects of regulation and road marking.

To solve this problem, a multiclass classification method is proposed, which is the next new and important result. For the multiclass classification, six classes were chosen: vehicles, people, road signs, traffic lights, road crossings, and road markings. We found that, despite the additional classes, the results of the HOG–BOVA–BEN method changed slightly. In the case of using the HOG–BoVW–BPNN classifier, it was possible to identify objects of different types with an accuracy of 71 to 86% (depending on the type), which shows the advantage of our approach compared to using the CNN method. Without a doubt, the proposed classification is not exhaustive and requires additional research. However, we consider the results obtained by the HOG–BoVW–BPNN method on the images of objects damaged by water drops and dirt to be very encouraging. The solution may be to use multiple cameras with overlapping viewing angles, which should greatly improve the object classification accuracy.

6. Conclusions

In our work, we evaluated deep learning methods for detecting objects in poor visibility conditions for video recorders. The images from the video cameras were taken through a protective glass covered with drops of water or dirt and were used as the objects of the study. Previously, the images were processed using the Canny operator, which allowed for the disposal of the damaged image blocks. The remaining blocks were examined by two HOG–BOW methods: a BPN and a CNN. Binary (machine–human) and multiclass classifications were performed. In the binary classification, the advantage of the HOG–BOVA–BEN method was established (accuracies of 79 vs. 65%, respectively). It was found that multiclass classification allows for the identification of objects with an accuracy of 71 to 86%. In order to increase the accuracy, the authors suggest using several cameras that overlap with each other’s viewing angles.

We plan to research to further improve the results of processing that is damaged by water drops and dirt images. In particular, we consider the use of polarimetry in combination with deep learning to be promising [47]. We hypothesize that polarimetry combined with deep learning can significantly improve the quality of object detection.

Author Contributions: Conceptualization, A.O. and E.P.; methodology, S.G.; software, S.K.; validation, M.I., A.F. and V.Y.; formal analysis, A.O.; investigation, E.P.; resources, E.P.; data curation, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: The results of part 2 “Theoretical background” and part 3 “Materials and Methods” were obtained within the Russian Science Foundation grant (project No. 20-71-10087). The authors of part 1 “Introduction”, part 4 “Experiments and Results”, part 5 “Discussion” and part 6 “Conclusions” are Aleksey Osipov, Ekaterina Pleshakova, Sergey Gataullin, Sergey Korchagin, Mikhail Ivanov (Financial University under the Government of the Russian Federation) and Vibhash Yadav (Rajkiya Engineering College).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Finogeev, A.; Parygin, D.; Schevchenko, S.; Finogeev, A.; Ather, D. Collection and Consolidation of Big Data for Proactive Monitoring of Critical Events at Infrastructure Facilities in an Urban Environment. In *Creativity in Intelligent Technologies and Data Science, Proceedings of the 4th International Conference CIT&DS 2021, Volgograd, Russia, 20–23 September 2021*; Springer: Cham, Switzerland, 2021; Volume 1448. [[CrossRef](#)]
2. Anokhin, A.; Burov, S.; Parygin, D.; Rent, V.; Sadovnikova, N.; Finogeev, A. Development of Scenarios for Modeling the Behavior of People in an Urban Environment. In *Society 5.0: Cyberspace for Advanced Human-Centered Society*; Studies in Systems, Decision and Control; Springer: Cham, Switzerland, 2021; Volume 333, pp. 103–114. [[CrossRef](#)]
3. Kolimenakis, A.; Solomou, A.D.; Proutsos, N.; Avramidou, E.V.; Korakaki, E.; Karetos, G.; Maroulis, G.; Papagiannis, E.; Tsagkari, K. The Socioeconomic Welfare of Urban Green Areas and Parks; A Literature Review of Available Evidence. *Sustainability* **2021**, *13*, 7863. [[CrossRef](#)]

4. Solomou, A.D.; Topalidou, E.T.; Germani, R.; Argiri, A.; Karetsos, G. Importance, utilization and health of urban forests: A review. *Not. Bot. Horti Agrobot. Cluj-Napoca* **2019**, *47*, 10–16. [\[CrossRef\]](#)
5. Grima, N.; Corcoran, W.; Hill-James, C.; Langton, B.; Sommer, H.; Fisher, B. The importance of urban natural areas and urban ecosystem services during the COVID-19 pandemic. *PLoS ONE* **2020**, *15*, e0243344. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Kondo, M.C.; Fluehr, J.M.; McKeon, T.; Branas, C.C. Urban green space and its impact on human health. *Int. J. Environ. Res. Public Health* **2018**, *15*, 445. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Braubach, M.; Egorov, A.; Mudu, P.; Wolf, T. Effects of urban green space on environmental health, equity and resilience. In *Nature-Based Solutions to Climate Change Adaptation in Urban Areas*; Springer: Cham, Switzerland, 2017; pp. 187–205.
8. Chiesura, A. The role of urban parks for the sustainable city. *Landsc. Urban Plan.* **2004**, *68*, 129–138. [\[CrossRef\]](#)
9. Chattopadhyay, D.; Rasheed, S.; Yan, L.; Lopez, A.A.; Farmer, J.; Brown, D.E. Machine Learning for Real-Time Vehicle Detection in All-Electronic Tolling System. In Proceedings of the 2020 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 24 April 2020; pp. 1–6. [\[CrossRef\]](#)
10. Deng, C.-X.; Wang, G.-B.; Yang, X.-R. Image edge detection algorithm based on improved Canny operator. In Proceedings of the 2013 International Conference on Wavelet Analysis and Pattern Recognition, Tianjin, China, 14–17 July 2013; pp. 168–172. [\[CrossRef\]](#)
11. Krakhmalev, O.; Korchagin, S.; Pleshakova, E.; Nikitin, P.; Tsibizova, O.; Sycheva, I.; Liang, K.; Serdechnyy, D.; Gataullin, S.; Krakhmalev, N. Parallel Computational Algorithm for Object-Oriented Modeling of Manipulation Robots. *Mathematics* **2021**, *9*, 2886. [\[CrossRef\]](#)
12. Shalev-Shwartz, S.; Shammah, S.; Shashua, A. On a formal model of safe and scalable self-driving cars. *arXiv* **2017**, arXiv:1708.06374.
13. Mancini, M.; Costante, G.; Valigi, P.; Ciarfuglia, T.A. Fast robust monocular depth estimation for obstacle detection with fully convolutional networks. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 4296–4303.
14. Jia, B.; Feng, W.; Zhu, M. Obstacle detection in single images with deep neural networks. *Signal Image Video Process* **2015**, *10*, 1033–1040. [\[CrossRef\]](#)
15. Wang, Z.; Liu, K.; Li, J.; Zhu, Y.; Zhang, Y. Various Frameworks and Libraries of Machine Learning and Deep Learning: A Survey. *Arch. Comput. Methods Eng.* **2019**, 1–24. [\[CrossRef\]](#)
16. Ni, J.; Chen, Y.; Chen, Y.; Zhu, J.; Ali, D.; Cao, W. A survey on theories and applications for self-driving cars based on deep learning methods. *Appl. Sci.* **2020**, *10*, 2749. [\[CrossRef\]](#)
17. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 847–858. [\[CrossRef\]](#)
18. Akter, R.; Hosen, I. CNN-based Leaf Image Classification for Bangladeshi Medicinal Plant Recognition. In Proceedings of the 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), Bangladesh, 21–22 December 2020; pp. 1–6.
19. Marino, S.; Beausery, P.; Smolarz, A. Weakly-supervised learning approach for potato defects segmentation. *Eng. Appl. Artif. Intell.* **2019**, *85*, 337–346. [\[CrossRef\]](#)
20. Afonso, M.; Blok, P.M.; Polder, G.; van der Wolf, J.M.; Kamp, J. Blackleg Detection in Potato Plants using Convolutional Neural Networks. *IFAC-PapersOnLine* **2019**, *52*, 6–11. [\[CrossRef\]](#)
21. Wu, A.; Zhu, J.; Ren, T. Detection of apple defect using laser-induced light backscattering imaging and convolutional neural network. *Comput. Electr. Eng.* **2020**, *81*, 106454. [\[CrossRef\]](#)
22. Kuznetsova, A.; Maleva, T.; Soloviev, V. Using YOLOv3 Algorithm with Pre- and Post-Processing for Apple Detection in Fruit-Harvesting Robot. *Agronomy* **2020**, *10*, 1016. [\[CrossRef\]](#)
23. Korchagin, S.; Serdechny, D.; Kim, R.; Terin, D.; Bey, M. The use of machine learning methods in the diagnosis of diseases of crops. *E3S Web Conf.* **2020**, *176*, 04011. [\[CrossRef\]](#)
24. Marino, S.; Beausery, P.; Smolarz, A. Unsupervised adversarial deep domain adaptation method for potato defects classification. *Comput. Electron. Agric.* **2020**, *174*, 105501. [\[CrossRef\]](#)
25. Puno, J.C.V.; Billones, R.K.D.; Bandala, A.A.; Dadios, E.P.; Calilune, E.J.; Joaquin, A.C. Quality Assessment of Mangoes using Convolutional Neural Network. In Proceedings of the 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), Bangkok, Thailand, 18–20 November 2019; pp. 491–495.
26. Sharma, D.K.; Malikov, V.; Parygin, D.; Golubev, A.; Lozhenitsina, A.; Sadovnikov, N. GPU-Card Performance Research in Satellite Imagery Classification Problems Using Machine Learning. *Procedia Comput. Sci.* **2020**, *178*, 55–64. [\[CrossRef\]](#)
27. Yin, H.; Gong, Y.; Qiu, G. Fast and efficient implementation of image filtering using a side window convolutional neural network. *Signal Process.* **2020**, *176*, 107717. [\[CrossRef\]](#)
28. Maksimovic, V.; Petrovic, M.; Savic, D.; Jaksic, B.; Spalevic, P. New approach of estimating edge detection threshold and application of adaptive detector depending on image complexity. *Optik* **2021**, *238*, 166476. [\[CrossRef\]](#)
29. Pawar, K.B.; Nalbalwar, S.L. Distributed canny edge detection algorithm using morphological filter. In Proceedings of the IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), Bangalore, India, 20–21 May 2016; pp. 1523–1527.

30. Dinesh Kumar, M.; Babaie, M.; Zhu, S.; Kalra, S.; Tizhoosh, H.R. A comparative study of CNN, BoVW and LBP for classification of histopathological images. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–7. [\[CrossRef\]](#)
31. Tseng, D.-C.; Wei, R.-Y.; Lu, C.-T.; Wang, L.-L. Image restoration using hybrid features improvement on morphological component analysis. *J. Electron. Sci. Technol.* **2019**, *17*, 100014. [\[CrossRef\]](#)
32. Sharifrazi, D.; Alizadehsani, R.; Roshanzamir, M.; Joloudari, J.H.; Shoeibi, A.; Jafari, M.; Hussain, S.; Sani, Z.A.; Hasanzadeh, F.; Khozeimeh, F.; et al. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomed. Signal Process. Control* **2021**, *68*, 102622. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Ravivarma, G.; Gavaskar, K.; Malathi, D.; Asha, K.; Ashok, B.; Aarthi, S. Implementation of Sobel operator based image edge detection on FPGA. *Mater. Today Proc.* **2021**, *45 Pt 2*, 2401–2407. [\[CrossRef\]](#)
34. Andriyanov, N.A.; Dementiev, V.E.; Tashlinskiy, A.G. Detection of objects in the images: From likelihood relationships toward scalable and efficient neural networks. *J. Comput. Opt.* **2022**, *46*, 139–159. [\[CrossRef\]](#)
35. Andriyanov, N.; Khasanshin, I.; Utkin, D.; Gataullin, T.; Ignar, S.; Shumaev, V.; Soloviev, V. Intelligent System for Estimation of the Spatial Position of Apples Based on YOLOv3 and Real Sense Depth Camera D415. *Symmetry* **2022**, *14*, 148. [\[CrossRef\]](#)
36. Sebyakin, A.; Soloviev, V.; Zolotaryuk, A. Spatio-Temporal Deepfake Detection with Deep Neural Networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). In Proceedings of the LNCS, 16th International Conference on Diversity, Divergence, Dialogue, iConference 2021, Beijing, China, 17–31 March 2021.
37. Pavlyutin, M.; Samoyavcheva, M.; Kochkarov, R.; Pleshakova, E.; Korchagin, S.; Gataullin, T.; Nikitin, P.; Hidirova, M. COVID-19 Spread Forecasting, Mathematical Methods vs. Machine Learning, Moscow Case. *Mathematics* **2022**, *10*, 195. [\[CrossRef\]](#)
38. Imani, E.; Javidi, M.; Pourreza, H.-R. Improvement of retinal blood vessel detection using morphological component analysis. *Comput. Methods Programs Biomed.* **2015**, *118*, 263–279. [\[CrossRef\]](#)
39. Kang, S.; Iwana, B.K.; Uchida, S. Complex image processing with less data—Document image binarization by integrating multiple pre-trained U-Net modules. *Pattern Recognit.* **2021**, *109*, 107577. [\[CrossRef\]](#)
40. Pratikakis, I.; Zagori, K.; Kaddas, P.; Gatos, B. ICFHR 2018 competition on handwritten document image binarization (H-DIBCO 2018). In Proceedings of the 2018 IEEE International Conference on Frontiers in Handwriting Recognition, Niagara Falls, NY, USA, 5–8 August 2018; pp. 489–493. [\[CrossRef\]](#)
41. Manoharan, S. An improved safety algorithm for artificial intelligence enabled processors in self driving cars. *J. Artif. Intell. Capsul. Netw.* **2019**, *1*, 95–104. [\[CrossRef\]](#)
42. Wu, D.; Xu, L.; Wei, T.; Qian, Z.; Cheng, C.; Guoyi, Z.; Hailong, Z. Research of Multi-dimensional Improved Canny Algorithm in 5G Smart Grid Image Intelligent Recognition and Monitoring Application. In Proceedings of the 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), Chengdu, China, 23–26 April 2021; pp. 400–404.
43. Yang, Y.; Zhao, X.; Huang, M.; Wang, X.; Zhu, Q. Multispectral image based germination detection of potato by using supervised multiple threshold segmentation model and Canny edge detector. *Comput. Electron. Agric.* **2021**, *182*, 106041. [\[CrossRef\]](#)
44. Yuan, L.; Xu, X. Adaptive Image Edge Detection Algorithm Based on Canny Operator. In Proceedings of the 4th International Conference on Advanced Information Technology and Sensor Application (AITS), Harbin, China, 21–23 August 2015; pp. 28–31. [\[CrossRef\]](#)
45. Xin, G.; Ke, C.; Xiaoguang, H. An improved Canny edge detection algorithm for color image. In Proceedings of the IEEE 10th International Conference on Industrial Informatics, Beijing, China, 25–27 July 2012; pp. 113–117. [\[CrossRef\]](#)
46. Boonarchatong, C.; Ketcham, M. Performance analysis of edge detection algorithms with THEOS satellite images. In Proceedings of the International Conference on Digital Arts Media and Technology (ICDAMT), Chiang Mai, Thailand, 1–4 March 2017; pp. 235–239.
47. Blin, R.; Ainouz, S.; Canu, S.; Meriaudeau, F. Road scenes analysis in adverse weather conditions by polarization-encoded images and adapted deep learning. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 27–32. [\[CrossRef\]](#)
48. Setiawan, B.D.; Rusydi, A.N.; Pradityo, K. Lake edge detection using Canny algorithm and Otsu thresholding. In Proceedings of the 2017 International Symposium on Geoinformatics (ISyG), Malang, Indonesia, 24–25 November 2017; pp. 72–76.
49. Gunawan, T.S.; Yaacob, I.Z.; Kartiwi, M.; Ismail, N.; Za'bah, N.F.; Mansor, H. Artificial neural network based fast edge detection algorithm for MRI medical images. *Indones. J. Electr. Eng. Comput. Sci.* **2017**, *7*, 123–130. [\[CrossRef\]](#)
50. Parthasarathy, G.; Ramanathan, L.; Anitha, K.; Justindhas, Y. Predicting Source and Age of Brain Tumor Using Canny Edge Detection Algorithm and Threshold Technique. *Asian Pac. J. Cancer Prev.* **2019**, *20*, 1409.
51. Wu, G.; Yang, D.; Chang, C.; Yin, L.; Luo, B.; Guo, H. Optimizations of Canny Edge Detection in Ghost Imaging. *J. Korean Phys. Soc.* **2019**, *75*, 223–228. [\[CrossRef\]](#)
52. Johari, N.; Singh, N. Bone fracture detection using edge detection technique. In *Soft Computing: Theories and Applications*; Springer: Singapore, 2018; pp. 11–19.
53. Kalbasi, M.; Nikmehr, H. Noise-Robust, Reconfigurable Canny Edge Detection and its Hardware Realization. *IEEE Access* **2020**, *8*, 39934–39945. [\[CrossRef\]](#)
54. Ahmed, A.S. Comparative study among Sobel, Prewitt and Canny edge detection operators used in image processing. *J. Theor. Appl. Inf. Technol.* **2018**, *96*, 6517–6525.

55. Xiao, Z.; Zou, Y.; Wang, Z. An improved dynamic double threshold Canny edge detection algorithm. In *MIPPR 2019: Pattern Recognition and Computer Vision, Proceedings of the Eleventh International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2019), Wuhan, China, 2–3 November 2019*; SPIE: Bellingham, WA, USA, 2020; Volume 11430, p. 1143016.
56. Wu, F.; Zhu, C.; Xu, J.; Bhatt, M.W.; Sharma, A. Research on image text recognition based on canny edge detection algorithm and k-means algorithm. *Int. J. Syst. Assur. Eng. Manag.* **2021**, 1–9. [\[CrossRef\]](#)
57. Lynn, N.D.; Sourav, A.I.; Santoso, A.J. Implementation of Real-Time Edge Detection Using Canny and Sobel Algorithms. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, 1096, 012079. [\[CrossRef\]](#)
58. Rahman, M.A.; Amin, M.F.I.; Hamada, M. Edge Detection Technique by Histogram Processing with Canny Edge Detector. In *Proceedings of the 2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII), Kaohsiung, Taiwan, 21–23 August 2020*; pp. 128–131.
59. Rezaatofghi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019*; pp. 658–666.
60. Yu, J.; Xu, J.; Chen, Y.; Li, W.; Wang, Q.; Yoo, B.; Han, J.J. Learning Generalized Intersection Over Union for Dense Pixelwise Prediction. In *Proceedings of the International Conference on Machine Learning, Online. 18–24 July 2021*; pp. 12198–12207.
61. Berman, M.; Triki, A.R.; Blaschko, M.B. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018*; pp. 4413–4421.
62. Lin, K.; Zhao, H.; Lv, J.; Zhan, J.; Liu, X.; Chen, R.; Li, C.; Huang, Z. Face detection and segmentation with generalized intersection over union based on mask R-CNN. In *Proceedings of the International Conference on Brain Inspired Cognitive Systems, Guangzhou, China, 13–14 July 2019*; Springer: Cham, Switzerland, 2019; pp. 106–116.
63. Kamyshova, G.; Osipov, A.; Gataullin, S.; Korchagin, S.; Ignar, S.; Gataullin, T.; Terekhova, N.; Suvorov, S. Artificial neural networks and computer vision's based Phytoindication systems for variable rate irrigation improving. *IEEE Access* **2022**, 10, 8577–8589. [\[CrossRef\]](#)
64. Wu, S.; Yang, J.; Yu, H.; Gou, L.; Li, X. Gaussian Guided IoU: A Better Metric for Balanced Learning on Object Detection. *arXiv* **2021**, arXiv:2103.13613.
65. Farhadi, A.; Redmon, J. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
66. Bischke, B.; Helber, P.; Folz, J.; Borth, D.; Dengel, A. Multi-task learning for segmentation of building footprints with deep neural networks. In *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019*; pp. 1480–1484.
67. Abouzahir, S.; Sadik, M.; Sabir, E. Bag-of-visual-words-augmented Histogram of Oriented Gradients for efficient weed detection. *Biosyst. Eng.* **2021**, 202, 179–194. [\[CrossRef\]](#)
68. Jogin, M.; Madhulika, M.S.; Divya, G.D.; Meghana, R.K.; Apoorva, S. Feature extraction using convolution neural networks (CNN) and deep learning. In *Proceedings of the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 18–19 May 2018*; pp. 2319–2323.
69. Bykov, A.; Grecheneva, A.; Kuzichkin, O.; Surzhik, D.; Vasilyev, G.; Yerbayev, Y. Mathematical Description and Laboratory Study of Electrophysical Methods of Localization of Geodeformational Changes during the Control of the Railway Roadbed. *Mathematics* **2021**, 9, 3164. [\[CrossRef\]](#)
70. Nasiri, A.; Taheri-Garavand, A.; Zhang, Y.-D. Image-based deep learning automated sorting of date fruit. *Postharvest Biol. Technol.* **2019**, 153, 133–141. [\[CrossRef\]](#)
71. Wei, Y.; Tian, Q.; Guo, J.; Huang, W.; Cao, J. Multi-vehicle detection algorithm through combining Harr and HOG features. *Math. Comput. Simul.* **2019**, 155, 130–145. [\[CrossRef\]](#)
72. Chitra, J.; Muthulakshmi, K.; Devi, K.G.; Balasubramanian, K.; Chitral, L. Review on intelligent prediction transportation system for pedestrian crossing using machine learning. *Mater. Today Proc.* **2021**. [\[CrossRef\]](#)