algorithms

MDPI

Article

Combining Background Subtraction and Convolutional Neural Network for Anomaly Detection in Pumping-Unit Surveillance

Tianming Yu, Jianhua Yang and Wei Lu *

School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China; yxm2013@mail.dlut.edu.cn (T.Y.); jianhuay@dlut.edu.cn (J.Y.)

* Correspondence: luwei@dlut.edu.cn

Received: 17 April 2019; Accepted: 24 May 2019; Published: 29 May 2019

Abstract: Background subtraction plays a fundamental role for anomaly detection in video surveillance, which is able to tell where moving objects are in the video scene. Regrettably, the regular rotating pumping unit is treated as an abnormal object by the background-subtraction method in pumping-unit surveillance. As an excellent classifier, a deep convolutional neural network is able to tell what those objects are. Therefore, we combined background subtraction and a convolutional neural network to perform anomaly detection for pumping-unit surveillance. In the proposed method, background subtraction was applied to first extract moving objects. Then, a clustering method was adopted for extracting different object types that had more movement-foreground objects but fewer typical targets. Finally, nonpumping unit objects were identified as abnormal objects by the trained classification network. The experimental results demonstrate that the proposed method can detect abnormal objects in a pumping-unit scene with high accuracy.

Keywords: background subtraction; transfer learning; classification

1. Introduction

Anomaly detection in video surveillance has become a public focus. It is an unsupervised learning task that refers to the problem of identifying abnormal patterns or motions in video data [1–3]. One of the most effective and frequently used methods of anomaly detection is to adopt background-subtraction methods in video surveillance. Over the past couple of decades, diverse background-subtraction methods have been presented by researchers to identify foreground objects in the videos [4–6]. The main idea of the background-subtraction algorithm is to build a background model [7], compare the current frame against the background model, and then detect moving objects according to their differences. There are some representative methods. For instance, Stauffer and Grimson proposed a Gaussian mixture model (GMM) for background modeling in cases of dynamic scenes, illumination changes, shaking trees, and so on [8]. Makantasis et al. estimated the thermal responses of each pixel of thermal imagery as a mixture of Gaussians by a Bayesian approach [9]. Barnich et al. applied random aggregation to background extraction and proposed the ViBe (visual background extractor) method [10]. In building a samples-based estimation of the background and updating the background models, ViBe uses a novel random selection strategy that indicates that information between neighboring pixels can propagate [11,12]. Elgammal et al. presented a nonparametric method based on kernel-density estimation (KDE) [13]. In this method, it is not necessary to estimate the parameter because it depends on previously observed pixel values, and there is no need to store the complete data. KDE has been commonly applied to vision processing, especially in cases where the underlying density is unknown. Hofmann et al. proposed the pixel-based adaptive segmenter (PBAS) in 2012 [14].

This algorithm, a nonparametric model based on pixels, combines the advantages of ViBe while making some improvements. It has realized nonparameter moving-object detection, and it is robust to slow illumination variation. St-Charles et al. proposed self-balanced sensitivity segmenter (SuBSENSE), which uses the principle of sample consistency and a feedback mechanism, which means that this background model can adapt to the diversity of complex backgrounds [15].

These existing background-subtraction methods are used to detect foreground objects in many applications showing good performance. However, in pumping-unit surveillance, the rotating pumping unit is judged as a foreground object when a traditional background-subtraction method is used for anomaly detection. Because the traditional background-subtraction method cannot obviate the interference of a rotating pumping unit, this results in losing the purpose of anomaly monitoring in video surveillance. On the other hand, intelligent monitoring systems are capable to detect unknown object types or unusual scenarios, whereas traditional background-subtraction methods can only provide the regions of abnormal objects and not give their specific category. Thus, the regions of interest, which are extracted from the image background by background-subtraction methods, need further processing.

In recent years, deep learning has made remarkable achievements in the field of computer vision. Deep learning is widely used in image recognition, object detection and classification [16,17]. This has achieved state-of-the-art results in those fields. GoogLeNet [18] is a deep convolutional neural network (CNN) [19]-based system that has been used in object recognition.

In this paper, we combined background subtraction and a CNN for anomaly detection in pumping-unit surveillance. In the proposed method, the background-subtraction method is used to extract motion objects in scenes, and a CNN identifies motion objects. A large quantity of samples is needed to train a deep CNN, but in practical application, it is always hard to provide enough samples. Therefore, a pretrained fine-tuned CNN was used in the proposed method.

The rest of this paper is organized as follows. Section 2 gives a brief introduction of pumping-unit surveillance. Section 3 presents the details of the proposed method. Section 4 shows the experiments on surveillance videos of the pumping unit to verify the validity and feasibility of the proposed method. Finally, conclusions are given in Section 5.

2. Problem of Pumping-Unit Surveillance

When a background-subtraction method is used for abnormal detection in a pumping-unit scene, the rotating pumping unit is extracted as a foreground object. As shown in Figure 1, the pumping unit is also detected as a foreground object as the vehicle. It is worth noting that several parts of the pumping unit are detected as the foreground rather than the whole pumping unit. In a normal situation, the rotating pumping unit should not be regarded as an abnormal object. To detect abnormal scenarios, a scene with a moving pumping unit that should be regarded as part of the background. Therefore, simply using background subtraction is not suitable for abnormal detection in a pumping-unit scene. The problem of pumping-unit surveillance is to detect real abnormal objects, and recognize and classify the objects. Figure 2 shows the outline of pumping-unit surveillance. Pumping units, vehicles, and pedestrians in pumping-unit scenes should be correctly identified and classified.

(**a**) (**b**)

Figure 1. Anomaly detection of pumping unit by a background-subtraction method: (**a**) pumping-unit scene; (**b**) foreground objects.

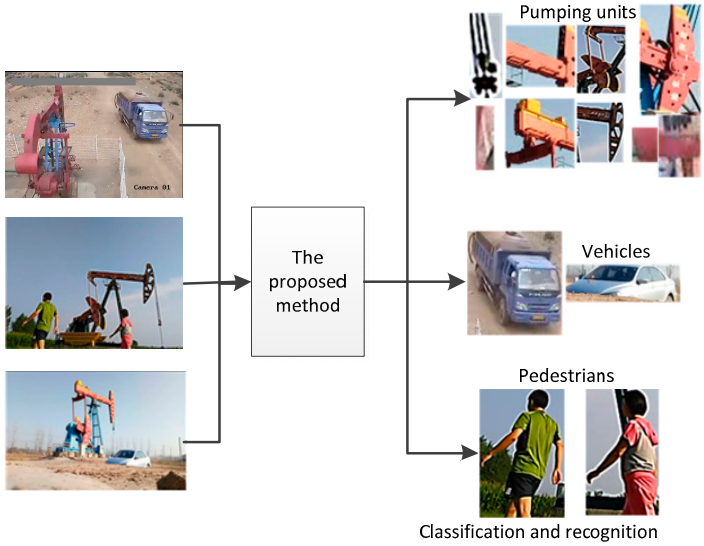

Figure 2. Outline of pumping unit surveillance.

3. Proposed Method

In this section, an intelligent method of pumping-unit surveillance is presented in detail. The system of pumping-unit surveillance is the centralized distributed architecture. Figure 3 shows the framework of the proposed method, including training and detection phases. In front-end processors, the input frame of each pumping-unit monitoring scene is processed by a background-subtraction method; so far, moving foreground objects are extracted. In a back-end processor, these objects are classified by clustering technology and then fed into the pretrained GoogLeNet [18]. Transfer learning method is used to retrain GoogLeNet. In this way, the classification network is completed, which is used for the classification and recognition of foreground objects.

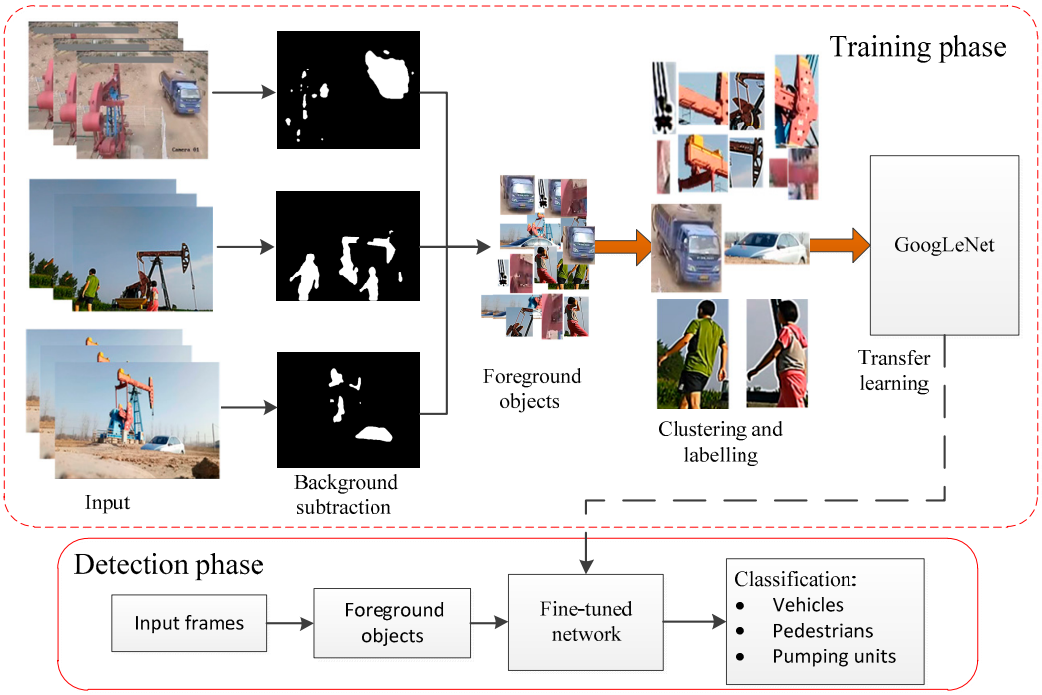

Figure 3. Framework of proposed method.

3.1. Moving-Object Extraction

Background subtraction is the basis of subsequent abnormal detection. In the training phase, the segmentation result obtained by background subtraction is used as a label mask. In the detecting phase, the foreground object obtained by background subtraction is used as the input of subsequent recognition and classification. In this way, it is only needed to judge and classify the foreground object rather than to recognize the whole image with a sliding window. Therefore, computation can be reduced and processing speed can be improved. The advantage of this method is that it is unsupervised; hence, the performance of background subtraction directly affects classification accuracy. In this paper, SuBSENSE [20], a state-of-the-art unsupervised background-subtraction method, was adopted for extracting the foreground object in the video. SuBSENSE is a pixel-level background-subtraction algorithm, and its basic idea is to use color and texture features to first detect moving objects, then introduce the idea of feedback control to adaptively update the parameters in the background model with the obtained rough segmentation results, so as to achieve better detection results. Foreground F can be obtained after video frame I is processed by SuBSENS:

$$F(i,j) = \begin{cases} 1, & if\ SuBSENSE(I(i,j))\ is\ foreground \\ 0, & if\ SuBSENSE(I(i,j))\ is\ background \end{cases} \quad (1)$$

where i and j are the position coordinates of the pixels. After obtaining the foreground pixels, the connected component-labeling method is used to locate and mark each connected region in the image, so as to obtain foreground target $\boldsymbol{O}$ [21]:

$$\boldsymbol{O} = blob(F) > n \quad (2)$$

where n is the least number of pixels in a connected region; in this paper, we set n = 150, namely, only the connected regions with more than 150 pixels were regarded as foreground objects.

3.2. Clustering and Labeling

In the training phase, a large number of moving objects $\boldsymbol{O}$ are extracted by the background-subtraction method. According to prior knowledge, these objects have two characteristics: (1) numerous objects; (2) fewer categories.

Several parts of the pumping unit are detected as foreground targets, which are classified into the same category. The moving objects that need to be recognized in the pumping-unit monitoring site are divided into three categories: pumping unit, vehicle, and pedestrian. There are many kinds of clustering algorithms that are used to deal with data-structure partition [22–24]. In this paper, foreground objects are subdivided into several subcategories by a hierarchical clustering algorithm, and then these subcategories are divided into pumping unit, vehicle, and pedestrian through human intervention, which are used as the training data of GoogLeNet.

Strategies for hierarchical clustering generally fall into two types, agglomerative and divisive [25]. This clustering method uses data linkage criteria to repeatedly merge or split the data to build a hierarchy of clusters through a hierarchical architecture. The clustering process is as follows:

(1) Assuming that foreground moving object $\boldsymbol{O} = \{o_1, o_2, \cdots o_k\}$ has k samples, the resolution of foreground moving object $\boldsymbol{O}$ is resized to 224×224.

(2) Samples are aggregated by a bottom–up approach, and Euclidean distance is chosen as the similarity measurement between categories:

$$d(o_i, o_j) = \left\| o_i - o_j \right\|_2 \tag{3}$$

where $i, j = 1, 2 \cdots k$. Linkage criteria use the average distance between all pairs of objects in any two clusters:

$$D(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d\left(o_{ri}, o_{sj}\right), \tag{4}$$

where r and s are clusters and n_r and n_s are the number of objects in cluster r and s, respectively. Similarly, o_{ri} and o_{sj} are the ith and jth object in cluster r and s, respectively.

(3) The pedestrian and vehicle categories in hierarchical clustering are selected separately, and the other categories are classified as part of the pumping-unit category.

Figure 4 shows the clustering process of foreground objects.

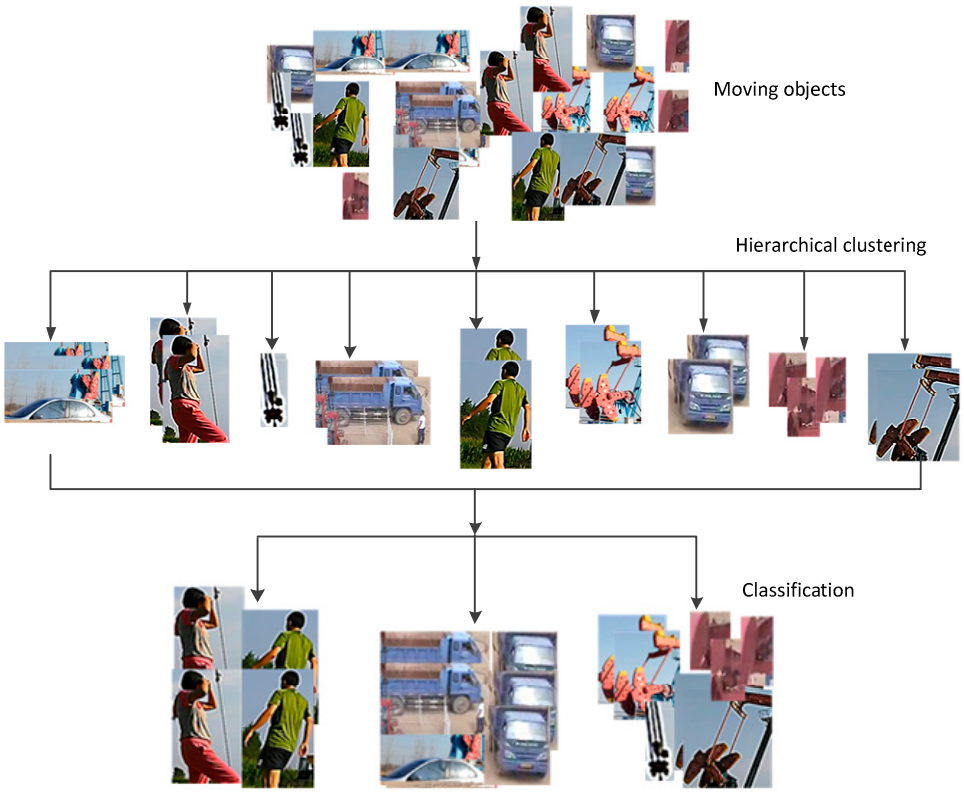

Figure 4. Clustering and labeling.

3.3. Transfer Learning

In traditional machine learning, a training set and test set are required to be in the same feature space and have the same data distribution. However, this demand is not satisfied in many cases, unless plenty of time and effort are spent to label the mass as of date. Transfer learning is a branch of machine learning. It can apply trained data to new problems, which can help avoid many data-labeling efforts. As deep learning develops quickly, transfer learning is increasingly combined with neural networks. In this paper, we used parameter-based transfer learning to address the problem of lacking abundant image samples of the labeled pumping unit.

In the classification application of pumping-unit monitoring, it is very time consuming to retrain a new neural network. Training data are not rich enough to train a deep neural network with strong generalization ability. To address this problem, transfer learning is desirable. For the past few years, transfer learning has been widely applied in various fields [26,27]. Pretrained models are usually based on large datasets, which can expand our training data, make the model more robust, improve the generalization ability, and save the time cost of training. The weight of the pretrained network is initialized and then fine-tuned on the new data. Compared with retraining the weight of network, this method can achieve better accuracy.

GoogLeNet is a pretrained convolutional neural network; it was trained on ImageNet [28], which has a million images. In this paper, GoogLeNet was retrained in pumping-unit data to classify objects that were extracted in the pumping-unit scene. Figure 5 shows the architecture of the fine-tuned GoogLeNet. Replacing the last three layers of GoogLeNet are a fully connected layer, a softmax layer, and a classification output layer. These three layers combine the general features of the objects extracted by the network, and convert the objects into the probability of different category labels. The size of the final full connection layer was set to 3, which is the same as the number of object categories in the pumping data. Then, the earlier layers in the network were frozen, that is, in subsequent training, the learning rate of these layers was set to 0, and the weight parameters of these layers were kept unchanged. Freezing earlier layers not only speeds up training, but also prevents overfitting of the

pumping data. In this paper, the layers before inception 5a were frozen, and the layers behind it were retrained. The loss function is cross-entropy loss, and the L_2 regularization term of the weights was added to the loss function to alleviate the effect of overfitting. Thus, the objective function was as follows:

$$w^* = arg\min_w \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n} t_{ij}\log y_{ij} + \frac{1}{2}\lambda w^T w \tag{5}$$

where m is the number of samples, n is the number of classes, t_{ij} is the indicator that the ith sample belongs to the jth class, w is the weight vector, and λ is the regularization factor. y_{ij} is the value from the softmax function, which is the output of sample i for class j:

$$y_{ij} = softmax\left(z_{ij}\right) = \frac{e^{z_{ij}}}{\sum_j e^{z_{ij}}} \tag{6}$$

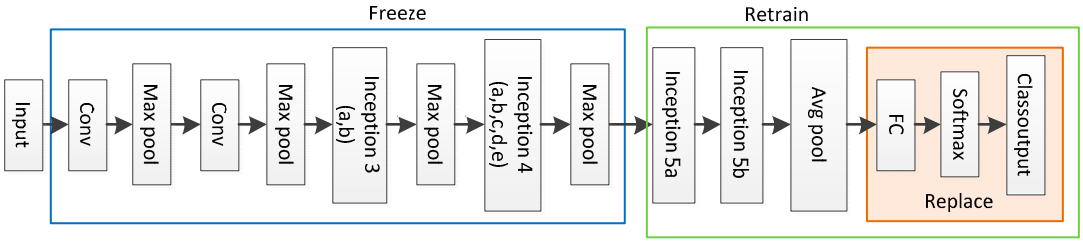

Figure 5. Fine-tuning GoogLeNet.

4. Experiments

In this section, four surveillance videos of pumping units were used to test the performance of the proposed method. Table 1 shows the details of these video datasets.

Table 1. Details of video datasets.

Data	Frame Dimension	FPS	Number of Frames	Objects
video 1	320 × 240	24	1677	Pumping unit, person
video 2	352 × 288	24	1708	Pumping unit, person, vehicle
video 3	640 × 480	24	1643	Pumping unit, person
video 4	640 × 480	24	4031	Pumping unit, person, vehicle

There are several performance indicators used to quantificationally evaluate the performance of the classification model [29]:

$$\begin{aligned} Accuracy &= \tfrac{TP+TN}{TP+TN+FP+FN}, \\ Recall &= \tfrac{TP}{TP+FN}, \\ Precision &= \tfrac{TP}{TP+FP}, \\ Specificity &= \tfrac{TN}{TN+FP}, \\ F_1 &= 2\times\tfrac{Precision\times Recall}{Precision+Recall}, \end{aligned}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. The higher the value of these indicators, the better the performance of the classification model.

4.1. Foreground Detection

The input video frame was segmented into foreground and background by the SuBSENSE algorithm, and multiple foreground objects were extracted. SuBSENSE [15,20] combines the color and local binary-similarity pattern features to detect moving objects. This method outperformed all previously tested state-of-the-art unsupervised methods on the CDnet [30] dataset. As a famous

benchmark dataset, CDnet provides ground truths for all video frames that range over diverse detection challenges such as dynamic background and various lighting conditions. Based on its excellent performance, SuBSENSE was used to extract the moving objects. Figure 6 presents the results of background subtraction. As can intuitively be seen, the segmentation results of SuBSENSE outperformed other methods. In foreground detection, several parts of the pumping unit were normally detected as the foreground rather than the whole pumping unit. The reason is that pumping units have a large scale along with periodic rotation in surveillance scenes. Some parts of the pumping unit are judged as background by background-subtraction methods.

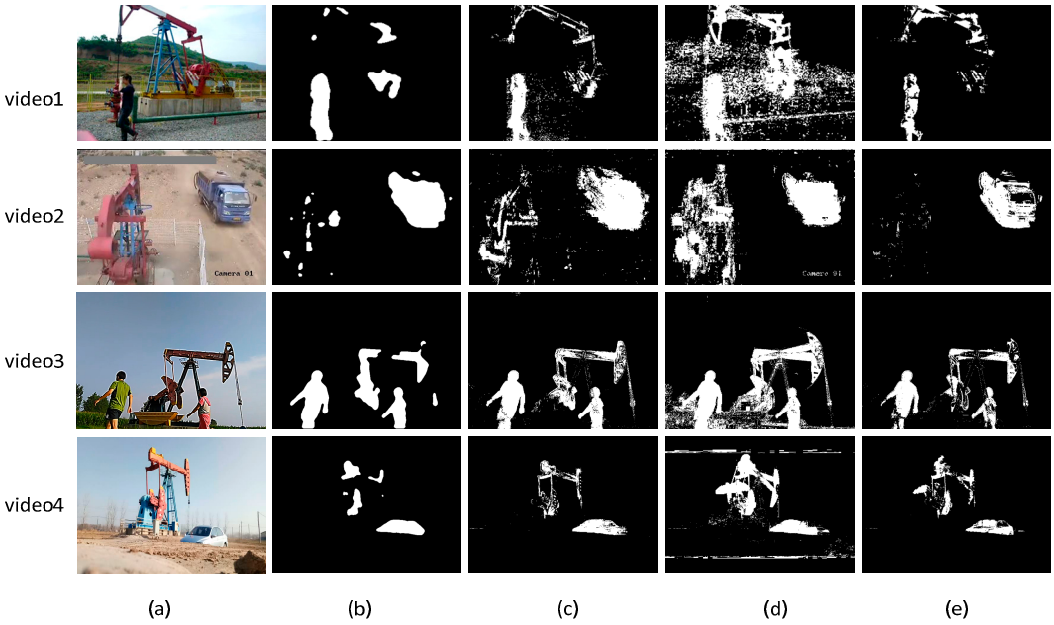

Figure 6. Comparisons of foreground-segmentation results. (**a**) Input images; (**b**) SuBSENSE; (**c**) Gaussian mixture model (GMM); (**d**) kernel-density estimation (KDE); (**e**) ViBe.

Pumping unit surveillance is a long time supervision; therefore, the background subtraction method has to address the light condition changes. In order to further verify the foreground extraction ability of the background subtraction method in light condition changes, a long term video was tested. Figure 7 shows the background subtraction results in the variant light conditions. As can intuitively be seen, the region of foreground detection of the pumping unit is less sensitive to the changing light. The experimental results show that SuBSENSE is able to eliminate the interference caused by light condition gradual changes.

Figure 7. Foreground detection in light condition changes cases. Screenshots and corresponding foreground detection results are illustrated from the first to the second rows, respectively. Numbers in the third row are time.

4.2. Object Classification

Through the clustering method mentioned in Section 3.2, these foreground objects were classified into three categories: pumping unit, person, and vehicle. In total, 1200 images were randomly selected as the image dataset to train and verify the performance of the classification network, which included 500 images of the pumping unit, 500 person images, and 200 vehicle images. In the monitoring video, there were a large number of foreground objects and a small number of typical targets, which means that each category of targets appeared repeatedly. 30 percent of images in the image dataset were randomly selected as the training set, and the remaining 70% as the testing set. The training process of the classification network is shown in Figure 8. The model tends to be convergent after 50 training iterations. The trained model can achieve high accuracy and low loss.

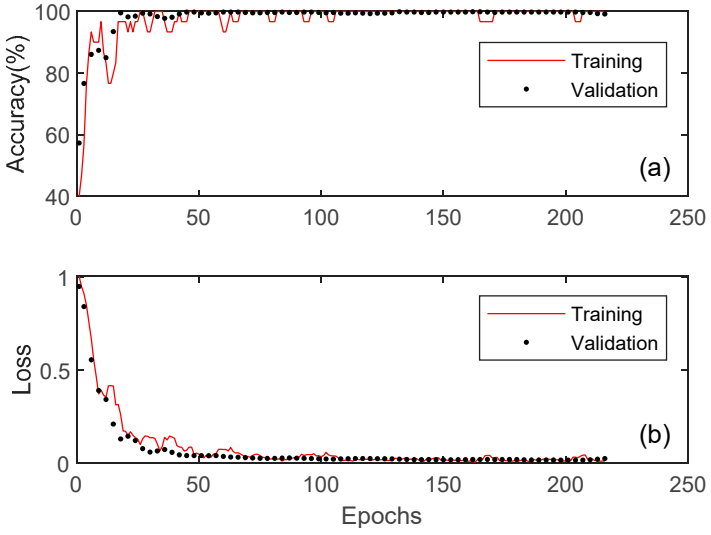

Figure 8. Training process of GoogLeNet. (**a**) Accuracy and (**b**) loss curves.

The classification network obtained by retraining GoogLeNet through the fine-tuned method was used for moving-object detection in the pumping-unit monitoring scene. Figure 9 shows the classifications of moving objects in the scene identified by the classification network. After moving objects are recognized and classified, the pumping unit is not regarded as an abnormal object, while persons and vehicles were output as abnormal objects. If there is no moving pumping unit in the

detected foreground objects, it means that the pumping unit has stopped working, and an abnormal alarm should be given.

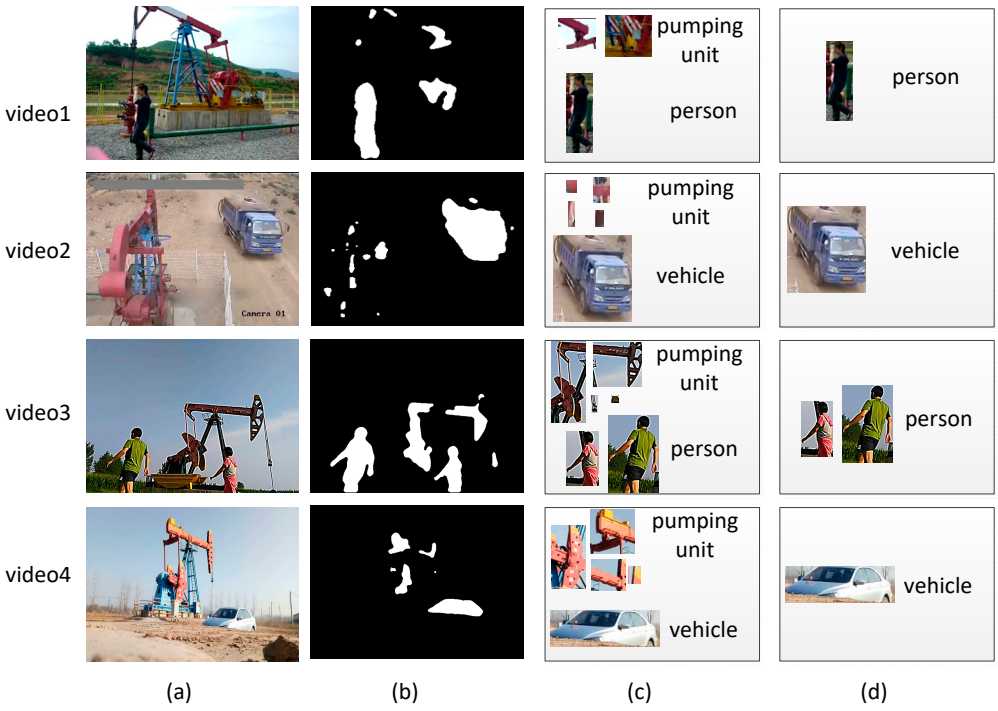

Figure 9. Classification of moving objects by retrained GoogLeNet. (**a**) Input images; (**b**) foreground; (**c**) classification; (**d**) anomaly objects.

To evaluate the proposed method, a histogram of oriented gradient (HOG) features and a multiclass support vector machine (SVM) classifier were used for comparative experiments. SVM is a classical classification method, while HOG features are a feature descriptor that is used for object detection in computer vision and image processing. It forms the features by calculation and statistics of the HOG in local areas of the image. HOG features combined with SVM classifiers have been widely used in image recognition [31]. The confusion matrices of the retrained net and SVM are presented in Figures 10 and 11, respectively. The experiment classification results of the three classes are listed in Table 2. To assure confidence in the experimental results, the experiment process was repeated 10 times. The average values of each metric are reported. The overall accuracy of the proposed method was 0.9988, while of the SVM was 0.9500. In the application of pumping-unit monitoring, the performance of the proposed method was obviously better than that ofa the classical SVM with HOG features.

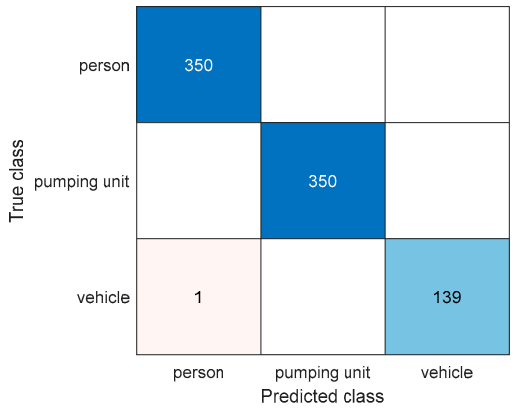

Figure 10. Confusion matrix of retrained GoogLeNet.

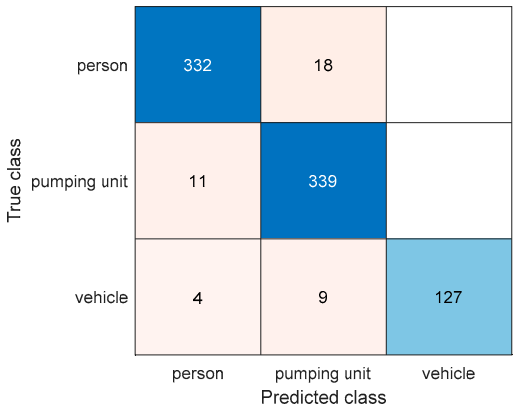

Figure 11. Confusion matrix of support vector machine (SVM).

Table 2. Experimental results.

Classes	Methods	Accuracy	Recall	Precision	Specificity	F_1
person	proposed	0.9988	1.0000	0.9972	0.9980	0.9986
	SVM	0.9607	0.9486	0.9568	0.9694	0.9527
pumping unit	proposed	1.0000	1.0000	1.0000	1.0000	1.0000
	SVM	0.9548	0.9686	0.9262	0.9449	0.9469
vehicle	proposed	0.9988	0.9929	1.0000	1.0000	0.9964
	SVM	0.9845	0.9071	1.0000	1.0000	0.9513

5. Conclusions

On-site monitoring of pumping units is a typical monitoring scene, that is, there is interference of periodic moving objects in the scene. The traditional background-subtraction method cannot satisfy the requirements of anomaly monitoring in this scenario. In the proposed method, background subtraction can extract possible abnormal targets. The pretrained CNN has a strong generalization and transplantation ability, which only needs a small number of samples and computing resources for retraining. After being trained by transfer learning, the network can be used to detect abnormal targets in a pumping-unit scene. The experimental results show that the proposed method can identify real foreground objects with high accuracy.

Author Contributions: Writing—original draft, T.Y.; Writing—review & editing, J.Y. and W.L.

Funding: This research was supported by the Natural Science Foundation of China (61876029).

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

References

1. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 15. [CrossRef]
2. Christiansen, P.; Nielsen, L.N.; Steen, K.A.; Jorgensen, R.N.; Karstoft, H. DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. *Sensors* **2016**, *16*, 1904. [CrossRef] [PubMed]
3. Kiran, B.R.; Thomas, D.M.; Parakkal, R. An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *J. Imaging* **2018**, *4*, 36. [CrossRef]
4. Brutzer, S.; Höferlin, B.; Heidemann, G. Evaluation of background subtraction techniques for video surveillance. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2011**, *32*, 1937–1944.
5. Toyama, K.; Krumm, J.; Brumitt, B.; Meyers, B. Wallflower: Principles and practice of background maintenance. *IEEE Int. Conf. Comput. Vis.* **1999**, *1*, 255–261.
6. Alan, M.M. Background subtraction techniques. *Proc. Image Vis. Comput.* **2000**, *2*, 1135–1140.

7. Babacan, S.D.; Pappas, T.N. Spatiotemporal algorithm for background subtraction. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP '07, Honolulu, HI, USA, 15–20 April 2007; pp. 1065–1068.
8. Stauffer, C.; Grimson, W.E.L. Adaptive background mixture models for real-time tracking. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **1999**, *2*, 246–252.
9. Makantasis, K.; Nikitakis, A.; Doulamis, A.D.; Doulamis, N.D.; Papaefstathiou, I. Data-driven background subtraction algorithm for in-camera acceleration in thermal imagery. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 2090–2104. [CrossRef]
10. Barnich, O.; Droogenbroeck, M.V. ViBe: A powerful random technique to estimate the background in video sequences. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 945–948.
11. Barnich, O.; Droogenbroeck, M.V. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.* **2011**, *20*, 1709–1724. [CrossRef]
12. Droogenbroeck, M.V.; Paquot, O. Background subtraction: Experiments and improvements for ViBe. *Comput. Vis. Pattern Recognit. Workshops* **2012**, *71*, 32–37.
13. Elgammal, A.; Harwood, D.; Davis, L. Non-parametric model for background subtraction. *Eur. Conf. Comput. Vis.* **2000**, *1843*, 751–767.
14. Hofmann, M.; Tiefenbacher, P.; Rigoll, G. Background segmentation with feedback: The pixel-based adaptive segmenter. In Proceedings of the IEEE Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16–21 June 2012; pp. 38–43.
15. St-Charles, P.-L.; Bilodeau, G.-A.; Bergevin, R. Flexible background subtraction with self-balanced local sensitivity. In Proceedings of the IEEE Computer Vision and Pattern Recognition Workshops, Montreal, QC, Canada, 23–28 June 2014; pp. 408–413.
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *1*, 1097–1105. [CrossRef]
17. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
18. Christian, C.; Liu, W.; Jia, Y.Q.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
19. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
20. St-Charles, P.-L.; Bilodeau, G.-A.; Bergevin, R. Subsense: A universal change detection method with local adaptive sensitivity. *IEEE Trans. Image Process.* **2015**, *24*, 359–373. [CrossRef] [PubMed]
21. Haralick, R.M.; Shapiro, L.G. *Computer and Robot Vision*; Addison-Wesley: Readimg, Boston, MA, USA, 1992; Volume 1, pp. 28–48.
22. Xu, D.; Tian, Y. A comprehensive survey of clustering algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [CrossRef]
23. Protopapadakis, E.; Voulodimos, A.; Doulamis, A.; Doulamis, N.; Dres, D.; Bimpas, M. Stacked autoencoders for outlier detection in over-the-horizon radar signals. *Comput. Intell. Neurosci.* **2017**. [CrossRef] [PubMed]
24. Protopapadakis, E.; Niklis, D.; Doumpos, M.; Doulamis, A.; Zopounidis, C. Sample selection algorithms for credit risk modelling through data mining techniques. *Int. J. Data Min. Model. Manag.* **2019**, *11*, 103–128. [CrossRef]
25. Lior, R.; Maimon, O. Clustering methods. In *Data Mining and Knowledge Discovery Handbook*; Springer: New York, NY, USA, 2005; pp. 321–352.
26. Patel, V.M.; Gopalan, R.; Li, R.; Chellappa, R. Visual domain adaptation: A survey of recent advances. *IEEE Signal Process. Mag.* **2015**, *32*, 53–69. [CrossRef]
27. Zhang, L. Transfer Adaptation Learning: A Decade Survey. *arXiv* **2019**, arXiv:1903.04687.
28. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis. (IJCV)* **2015**, *115*, 211–252. [CrossRef]
29. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.

30. Wang, Y.; Jodoin, P.-M.; Porikli, F.; Janusz, K.; Benezeth, Y.; Ishwar, P. CDnet 2014: An expanded change detection benchmark dataset. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 387–394.
31. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2005**, *1*, 886–893.