Computer Engineering and Applications Vol. 13, No. 1, February 2024

Video Anomaly Classification Using Convolutional Neural Network

Muhammad Naufal Rachmatullah[1)] Sutarno[2*)], Rahmat Fadli Isnanto[3)]

1) Department of Informatics, Faculty of Computer Science
2,3) Department of Computer Systems, Faculty of Computer Science
Sriwijaya University
Jl. Raya Palembang - Prabumulih No.KM. 32, Indonesia.

** sutarno@unsri.ac.id*

ABSTRACT

The use of surveillance videos is increasingly popular in city monitoring systems. Generally, the analysis process in surveillance videos still relies on conventional methods. This method requires professional personnel to constantly monitor and analyze videos to identify abnormal events. Consequently, the conventional approach is time-consuming, resource-intensive, and costly. Therefore, a system is needed to automatically detect video anomalies, reducing the massive human resource utilization for video monitoring. This research employs deep learning methods to classify anomalies in videos. The video anomaly detection process involves transforming the video into image format by extracting each frame present in the video. Subsequently, a Convolutional Neural Network (CNN) model is utilized to classify anomalous events within the video. Testing results using the CNN architectures DenseNet121 and EfficientNet V2 yielded performance accuracies of 99.89 and 98.24, respectively. The testing results indicate that the DenseNet121 architecture outperforms the EfficientNet V2 architecture in terms of performance.

Keywords: Video Anomaly Classification, Surveillance Video, DenseNet121, EfficientNetV2.

1. INTRODUCTION

The increasing security requirements in major cities have led to the widespread use of surveillance videos to monitor human activities and prevent unusual incidents such as traffic accidents and criminal violations [1]. Typically, the process of detecting video anomalies requires video analysis to distinguish abnormal actions. Many cities still employ conventional methods for surveillance video analysis, which necessitate constant monitoring and analysis by professional personnel, consuming significant time, effort, and cost [2].

Therefore, research activities on automatic video anomaly detection and analysis are necessary to improve service quality and transform surveillance systems into practical and significant ones. Effective detection techniques can reduce the human resources required for video monitoring, especially for surveillance systems requiring high accuracy levels. With automatic video anomaly detection, the system can pinpoint where anomalies occur in the video, expediting the assessment of unusual or suspicious activities. This system facilitates swift authorization related to identifying

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

the root causes of anomalies, thus saving time and effort needed for manual record search [1].

Anomalies in video data exhibit characteristics such as ambiguity, novelty, unfamiliarity, uncommonness, irregularity, unexpectedness, and unconventionality. These inherent characteristics pose challenges for modeling approaches in video anomaly detection [3]. The main challenges in video anomaly detection include inherent imbalance in video data between positive and negative classes, disruption of supervised learning-based algorithms due to high variation in positive classes, and poorly defined activities in video anomalies due to their high level of ambiguity [4 - 5].

Convolutional Neural Network (CNN) methods, known for their excellent performance on image datasets, are employed for video anomaly detection [3-9]. CNNs can effectively learn important features in image datasets, allowing them to yield good results for complex problems. In several studies on dominant anomaly detection in videos, CNN methods have demonstrated proficiency in identifying abnormal behavioral patterns. CNN methods optimize feature extraction processes in each video frame, leading to good classification results.

This study compares two CNN architectures for video anomaly detection: DenseNet121 and EfficientNetV2. DenseNet121 features inter-block connections, enhancing feature generation processes. On the other hand, EfficientNetV2 is renowned for its high accuracy levels in image classification with relatively fewer parameters. Both architectures leverage knowledge from pre-training on large datasets such as ImageNet, which comprises over 1.4 million images across more than 20,000 categories, during the model training process.

2. RESEARCH METHODOLOGY

The process of classifying anomalies in videos begins with data preparation as the initial step. In this research, data in the form of digital images in .jpg format is utilized, categorized into 13 anomaly classes and one normal class. Subsequently, the data undergoes preprocessing stages consisting of resizing and pixel normalization. The research then proceeds with training the DenseNet121 and EfficientNetV2 architectures. The model training process involves dividing the data into training and test sets. The two models resulting from the training phase are then evaluated using test data. The test metric results are analyzed and compared to determine the best CNN architecture for video anomaly classification. Generally, the steps of the video classification method using CNN are presented in Figure 1.

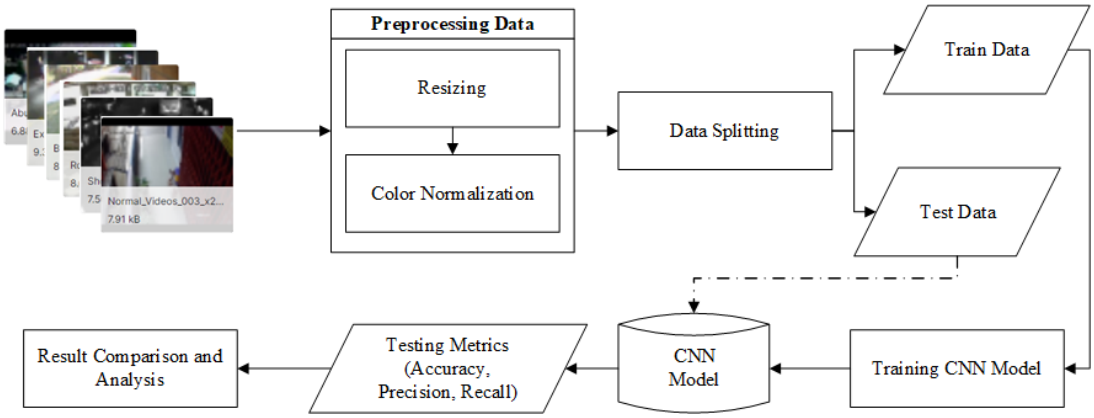

FIGURE 1. Video Anomaly Classification Methodology

2.1 DATASET

The dataset used in this study is the UCF Crime Dataset obtained from Kaggle [1]. This dataset comprises 1,266,345 training images and 111,308 testing images. The UCF Crime dataset consists of 14 classes, one normal class, and 13 abnormal classes (Abuse, Arrest, Arson, Assault, Burglary, Explosion, Fighting, Road Accidents, Robbery, Shooting, Shoplifting, Stealing, and Vandalism). Figure 2 illustrates some examples of the data used. Furthermore, the distribution of training and testing data is shown in Figures 3a and 3b.

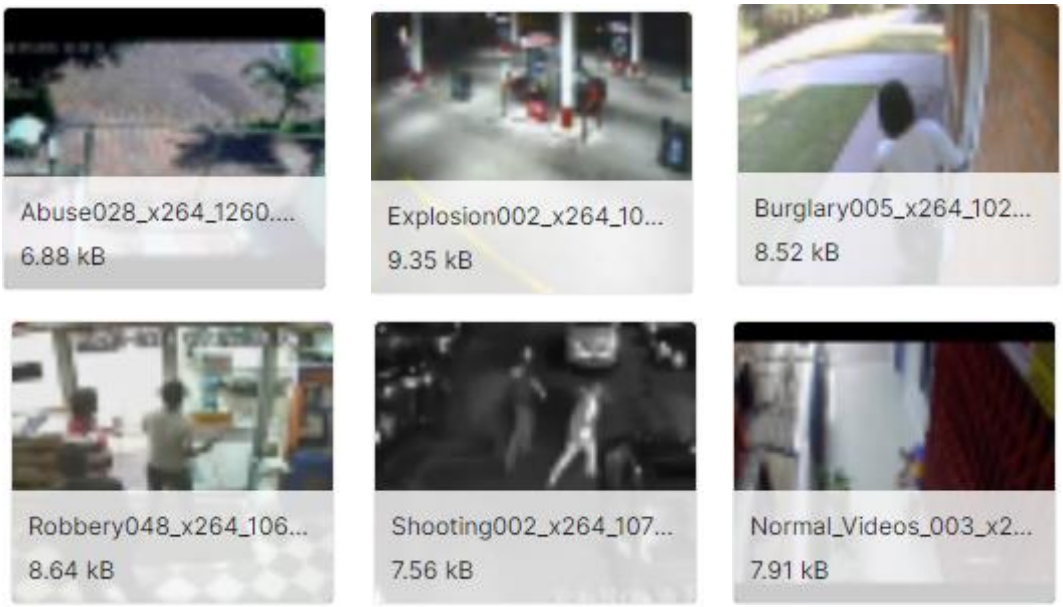

FIGURE 2. Example of the dataset used in this study

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

(a)

(b)

FIGURE 3. Data Distribution (a) Training (b) Testing

2.2 DenseNet121 Architecture

DenseNet is a model architecture with specific characteristics known as dense blocks, where each layer is directly connected to all other layers. A layer receives input from the output of all preceding layers and provides output to all subsequent layers, allowing the network to be more compact. This differs from traditional convolutional layers where a layer receives input from the preceding layer and provides output to the next layer. The Dense block is the part where the dimensions of the feature maps remain constant within a block, but the number of filters between them changes. In each block of the DenseNet architecture, a transition layer containing pooling layers is added. The Transition layers are used to reduce the number of features to half of the existing channels, enabling a reduction in computational overhead in this architecture.

2.3 EfficientNetV2 Architecture

EfficientNetV2 is an enhancement of the EfficientNet architecture designed to optimize computational efficiency and performance in CNNs. Developed by Google researchers, EfficientNetV2 introduces new design strategies to achieve superior results with fewer parameters. Unlike its predecessor, EfficientNetV2 utilizes a different scaling method called compound scaling. This method not only adjusts the width, depth, and resolution of the network but also introduces a new scaling dimension called "image size". This approach allows for more flexible model dimension adjustments, resulting in improved accuracy without excessive computation. Furthermore, EfficientNetV2 integrates advanced training techniques such as self-training and knowledge distillation to enhance model robustness and generalization capabilities. By leveraging state-of-the-art methods in model

architecture and training, EfficientNetV2 provides a significant improvement in CNN design, delivering superior performance in image classification cases while maintaining computational efficiency.

2.4 Evaluation Metrics

The evaluation process plays a crucial role in assessing the performance of CNN models in conducting anomaly video classification. This stage provides an analysis of the model's prediction accuracy regarding the categorization of input data. Evaluation metrics used to assess the classification model include accuracy, precision, recall, and F1-Score as seen in equations (1) – (4). To obtain the evaluation metric results of the model, a confusion matrix containing true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values is required.

$$Accuracy = \frac{(tp + tn)}{(tp + tn + fn + tn)} \quad (1)$$

$$Precision = \frac{(tp)}{(tp + fp)} \quad (2)$$

$$Recall = \frac{(tp)}{(tp + fn)} \quad (3)$$

$$F1Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} \quad (4)$$

3. RESULT AND DISCUSSIOIN

In the development of a classification model for video anomaly using CNN method, the test cases are divided into two types. The first test case involves performing multiclass classification on 13 classes of abnormalities (Abuse, Arrest, Arson, Assault, Burglary, Explosion, Fighting, Road Accidents, Robbery, Shooting, Shoplifting, Stealing, and Vandalism) and one normal class. Whereas, in the second test case, binary classification concept is employed by merging the 13 abnormal classes into one class, resulting in two classes: abnormal class, which is a combination of the 13 abnormal classes, and normal class. Both types of test cases undergo training and testing processes using two CNN architectures: DenseNet121 and EfficientNetV2.

3.1. EVALUATION RESULT ON TEST CASE 1

The evaluation process on the first test case begins with comparing the training results between the DenseNet121 and EfficientNetV2 architectures using a dropout function with a batch size of 32. This is done to assess the influence of dropout and batch size on the training results of the model. Furthermore, a dropout layer with a dropout ratio of 50% is inserted between the hidden layers and the output layer in the

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

fully connected operation. Table 1 presents a comparison of the evaluation results on the model trained using dropout and a batch size of 32.

TABLE 1.
Comparison Result on Testing Data Between DenseNet121 and EfficientNetV2

Class	Precision		Recall		F1-Score	
	EfficientNetV2	DenseNet121	EfficientNetV2	DenseNet121	EfficientNetV2	DenseNet121
Abuse	0	0	0	0	0	0
Arrest	2.38	3.57	13.55	4.71	4.05	4.06
Arson	29.0	35.84	28.92	30.12	28.98	32.73
Assault	0	5.66	0	24.59	0	9.20
Burglary	10.98	2.61	56	19.05	18.36	4.60
Explosion	0.92	2.92	10.90	42.22	1.69	5.46
Fighting	0.81	0.81	1.56	0.60	1.06	0.69
Normal	95.07	95.49	68.8	71.25	79.84	81.61
Road Accidents	28.57	31.96	11.93	21.52	16.83	25.72
Robbery	0	16.87	0	2.24	0	3.95
Shooting	0	1.57	0	20.69	0	2.92
Shoplifting	0.918	5.91	70	80.36	1.81	11.00
Stealing	19.1	4.04	19.89	5.00	19.53	4.47
Vandalism	0	0	0	0	0	0
Average	**13.42**	**14.80**	**20.11**	**23.02**	**12.29**	**13.32**

From Table 1, it is observed that the DenseNet121 and EfficientNetV2 models trained using dropout operation with a batch size of 32 yield an average performance below 50%. The highest classification performance values are obtained in the Normal class, with performance metrics of 95.49, 71.25, and 81.61 for precision, recall, and F1-Score, respectively, in the DenseNet121 architecture. The high performance of the model in the normal class is attributed to the significantly higher amount of data in that class compared to the other classes, leading to a phenomenon known as class imbalance. This is evident in the distribution of the model's prediction results shown in the confusion matrix (Figure 4). To mitigate the impact of class imbalance in the dataset, testing is conducted by merging all abnormal classes into one class, thus forming the second test case, which involves testing using two classes (normal and abnormal).

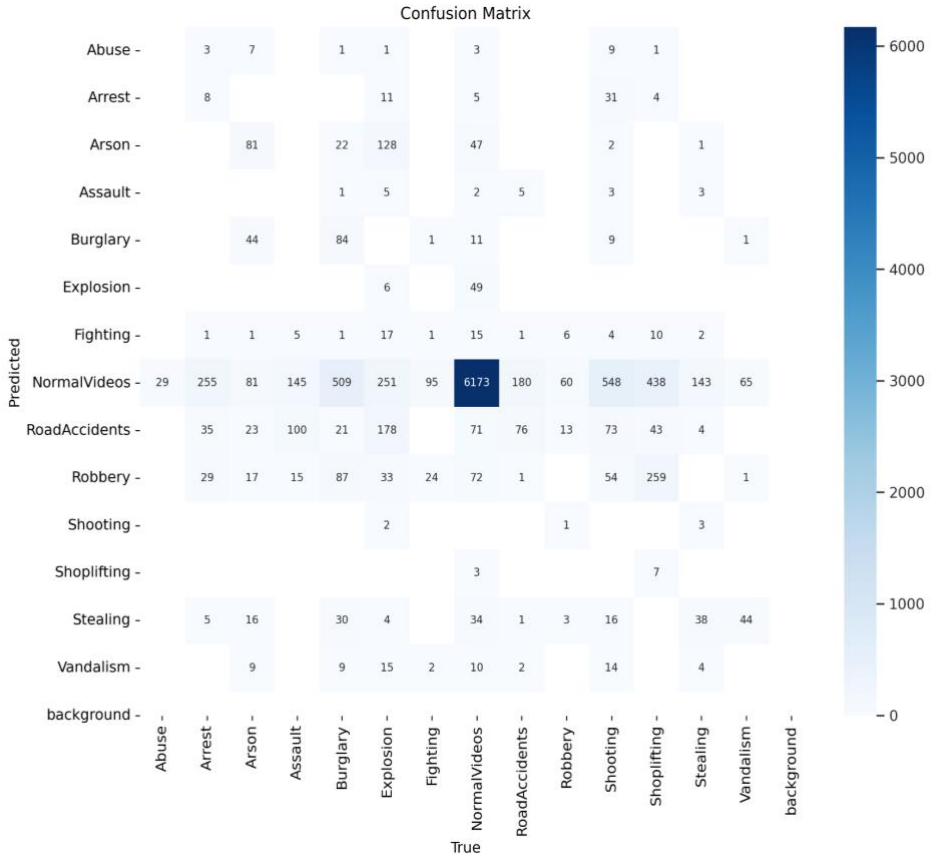

FIGURE4. Confusion Matrix of DenseNet121 on Testing Data

3.2. EVALUATION RESULT ON TEST CASE 2

In the second test case a binary classification scheme was done. Testing using the binary classification scheme involves the process of classifying data into two categories: normal and abnormal. In this context, all abnormal classes are merged into a single class to simplify the evaluation framework. This approach aims to address the issue of class imbalance observed in previous multiclass tests, where the normal class has a significantly larger number of samples than individual abnormal classes. Thus, binary classification testing enables a more balanced evaluation of the model's performance on both classes, without losing the importance of identifying various types of abnormalities. Through this approach, the model can be assessed more accurately in its ability to distinguish between normal situations and unusual or abnormal events in the context of video surveillance.

The results of the binary classification testing with DenseNet121 and EfficientNetV2 models reveal notable accuracy performances. Specifically, the DenseNet121 model achieved an accuracy of 99.89%, while the EfficientNetV2 model achieved a slightly lower accuracy of 98.24%. These accuracies signify the models' capabilities in effectively distinguishing between normal and abnormal instances within the video surveillance context. The higher accuracy achieved by the

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

DenseNet121 model underscores its robustness and efficacy in classification tasks compared to EfficientNetV2. These results indicate promising prospects for employing DenseNet121 in real-world applications where precise anomaly detection is paramount, while also highlighting the considerable competence of EfficientNetV2 in such tasks. Table 2 Illustrate the performance of DenseNet121 and EfficientNetV2 on testing data.

TABLE 2.
Evaluation Result in Binary Classification on DenseNet121 and EfficientNetV2

Class	Accuracy	Precision	Recall	F1-Score
DenseNet121	99.89	99.72	99.86	99.79
EfficientNetV2	98.24	98.92	98.92	98.92

4. CONCLUSION

From the test results, it can be concluded that the DenseNet121 model performed the best with very high levels of accuracy, precision, recall, and F1-score, reaching 99.89%, 99.72%, 99.86%, and 99.79% respectively. This outstanding performance indicates that DenseNet121 is a highly promising choice for classifying anomalies in videos. However, it is important to note that multi-class testing has certain drawbacks compared to binary class testing. In multi-class testing, there is a class imbalance issue where the normal class is significantly more represented than individual abnormal classes. This can lead to bias in model evaluation and difficulties in identifying anomalies in underrepresented classes. As a recommendation for further research, it is advisable to:

1. Further Research on Dataset Diversification: Conduct further research to collect a more balanced dataset between normal and abnormal classes, thereby reducing the class imbalance issue and improving the reliability of model evaluation.
2. Exploration of Data Imbalance Handling Techniques: Explore techniques such as oversampling, undersampling, or synthetic data generation to address data imbalances in minority classes, thus strengthening the model's ability to identify less common anomalies.
3. Development of Specialized Models for Multi-Class Cases: Develop specialized strategies and model architectures that can effectively handle multi-class cases by taking into account class imbalances and the complexity of distinguishing various types of anomalies.

By considering these recommendations, it is hoped that further research can enhance understanding and performance in video anomaly detection systems, as well as make valuable contributions to the development of security and monitoring technology.

ACKNOWLEDGEMENTS

"The research/publication of this article was Junded by DIPA oJ' Public Service Agency of Llniversitas Sriwijalta 2022- sp otpA-023.17.2.677515 /2022, On Desember 13, 2021. In accordance with the Rector's Decree Number: 0019/UN9/SK.LP2M.PT 12022, On Juni 15, 2022'.

REFERENCES

[1] V. Singh, S. Singh, and P. Gupta, "Real-Time Anomaly Recognition Through CCTV Using Neural Networks," Procedia Comput. Sci., vol. 173, no. 2019, pp. 254–263, 2020.
[2] K. Doshi and Y. Yilmaz, "Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate," Pattern Recognit., vol. 114, p. 107865, 2021.
[3] J. Arunnehru, "Deep learning-based real-world object detection and improved anomaly detection for surveillance videos," Mater. Today Proc., no. xxxx, 2021.
[4] Y. Zhou, H. Ren, Z. Li, and W. Pedrycz, "An anomaly detection framework for time series data : An interval-based approach," Knowledge-Based Syst., vol. 228, p. 107153, 2021.
[5] Alex Bewley;, Z. Ge;, L. Ott;, F. Ramos;, and B. Upcrof, "Simple online and realtime tracking," in 2016 IEEE International Conference on Image Processing (ICIP), 2016.
[6] K. K. Santhosh, D. P. Dogra, and P. P. Roy, "Anomaly Detection in Road Traffic Using Visual Surveillance : A Survey," vol. 53, no. 6, 2020.
[7] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2016, pp. 733–742, 2016.
[8] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection," Image Vis. Comput., vol. 106, p. 104078, 2021.
[9] Gao Huang and Zhuang Liu and Laurens van der Maaten and Kilian Q. Weinberger, Densely Connected Convolutional Networks, arXiv 1608.06993 (2016)

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)