ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 9 (2022) No.2, pp. 196-209
DOI:10.33168/LISS.2022.0212

A Study on Abnormal Behavior Detection in CCTV Images through the Supervised Learning Model of Deep Learning

Yangsun Lee

Seokyeong University, Seoul, Republic of Korea

yslee@skuniv.ac.kr

Abstract. Abnormal behavior refers to behavior that is atypical or statistically uncommon within a particular culture or that is maladaptive or detrimental to an individual or to those around that individual. Most of these actions are criminal acts that people fear. The installation of public CCTVs to prevent these crimes is increasing, but the crime rate is rather increasing recently. In line with this situation, artificial intelligence research using deep learning that automatically finds abnormal behavior in CCTV is increasing. Deep learning is a type of artificial intelligence designed based on artificial neural networks. C3D is a model that uses 3D Convolution Networks to train large amounts of image data in a supervised learning manner. In this paper, a study was conducted to learn and detect abnormal behavior images consisting of five crime types and 32 detailed types using the C3D model. As a result of learning and classifying abnormal behavior images with the C3D model, which is an artificial intelligence model, abnormal behavior was detected in several crime type image data, showing considerably high accuracy. Therefore, it is believed that abnormal behavior detection technology using C3D can contribute to protecting victims by classifying the precursor symptoms of crime by applying it to various CCTV image data to prevent crime in advance or detect criminal behavior.

Keywords: Abnormal behaviour, supervised learning model, deep learning, abnormal behaviour detection, C3D, 3D Convolution Networks

1. Introduction

Abnormal behavior refers to behavior that is not socially valuable or suitable for daily life. Most of these actions are criminal acts that people fear. The installation of public CCTVs to prevent these crimes is increasing, but the crime rate is rather increasing recently. In line with this situation, artificial intelligence research using deep learning that automatically finds abnormal behavior in CCTV is increasing. Deep learning is a type of artificial intelligence designed based on artificial neural networks. Convolutional 3D (C3D) is a model that uses 3D Convolutional Networks to train large amounts of image data in a supervised learning manner, and is effective in learning and classifying abnormal behavior images. With the development of human action recognition (HAR) abnormality detection research using machine learning and deep learning, cameras and CCTVs can recognize people and actions and classify objects, not just for filming or showing images. (Chalapathy, Chawla 2019. Chandola et al., 2009, Lee 2020, Park 2013).

In this paper, a study was conducted to learn and detect abnormal behavior in CCTV images using the C3D model, an artificial intelligence model. As a result, abnormal behavior was detected in CCTV image data of various crime types, showing quite high accuracy. Therefore, it is believed that the abnormal behavior detection technology using C3D can be applied to various CCTVs to prevent crimes in advance or to protect victims by detecting criminal behavior (Haroon, 2022, Ji et al., 2013, Kiran et al., 2018).

2. Related Studies

2.1. Abnormal behavior detection

Abnormal behaviour refers to behavior that is not socially valuable or suitable for daily life. Abnormal behavior Detection is a series of activities that find unexpected patterns in the data that learn the abnormal behavior and detect the abnormal behavior with the learning content.

Abnormal behavior detection is divided into three types according to learning data and circumstances. Supervised Anomaly Detection uses both labeled normal and abnormal data during learning, making model performance evaluation intuitive. Semi-supervised anomaly detection is useful in situations where only normal data is secured using only normal data. Unsupervised anomaly detection is mainly used when there is no labeling of data by assuming that most of the data is normal when learning (Chalapathy, Chawla 2019, Chandola et al., 2009, Lee, Lee 2021, Park 2021).

2.2. C3D Model

The C3D model is a model that uses 3D convolution networks to train large amounts of image data in a supervised learning manner. In addition, it has a 3x3x3 kernel

structure and spatio-temporal nature, showing great efficiency in learning image data.

In the past, image analysis was performed by 2D convolution networks. However, since 2D convolution networks had only spatial characteristics, it lost its temporal character, and the learning performance was poor because the result was in the form of an image. By compensating for these shortcomings, 3D convolution networks with spatial and temporal characteristics and output volume was created (Haroon 2022, Ji et al., 2013, Leng et al., 2019, Maturana, Scherer 2015, Tran et al., 2015, Wang et al., 2018, WU, Lee et al., 2019, Zeng 2019) . Figure 1 shows an example of the structure of C3D.

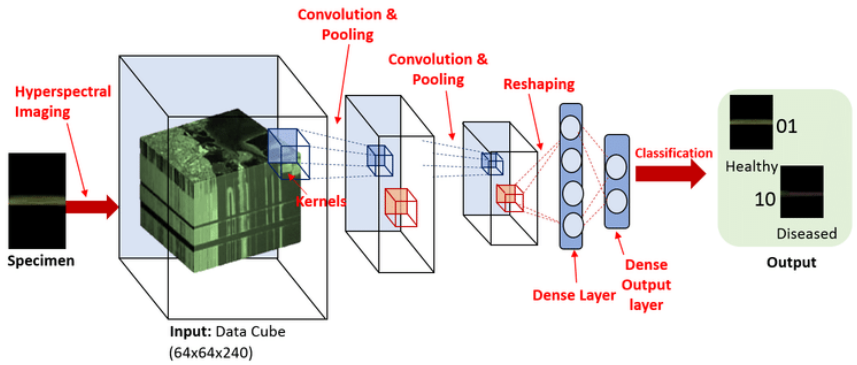

Figure 1. Example of 3D Convolution Networks

3. Learning and Detection of Abnormal Behavior in CCTV Images Using C3D

The C3D model using 3D Convolution Networks has a spatial and temporal nature, and shows great efficiency in learning abnormal behavior image data. In order to detect abnormal behavior in CCTV images using such C3D, it is first necessary to learn the abnormal behavior in CCTV images.

Figure 2 shows the process of learning and detecting abnormal behavior in CCTV images using the C3D model. First, a learning dataset is required for learning. Dataset construction generates a list after classification into learning and test images. An index item of an image is generated and an image is imaged. We then learn using the generated learning dataset and C3D model, and generate a weight file containing weights generated through learning for future anomaly behavior detection. The type is detected using the generated weight file and the abnormal behavior image to be tested. By evaluating abnormal behavioral images, scores are assigned to nearby types to derive result types and accuracy.

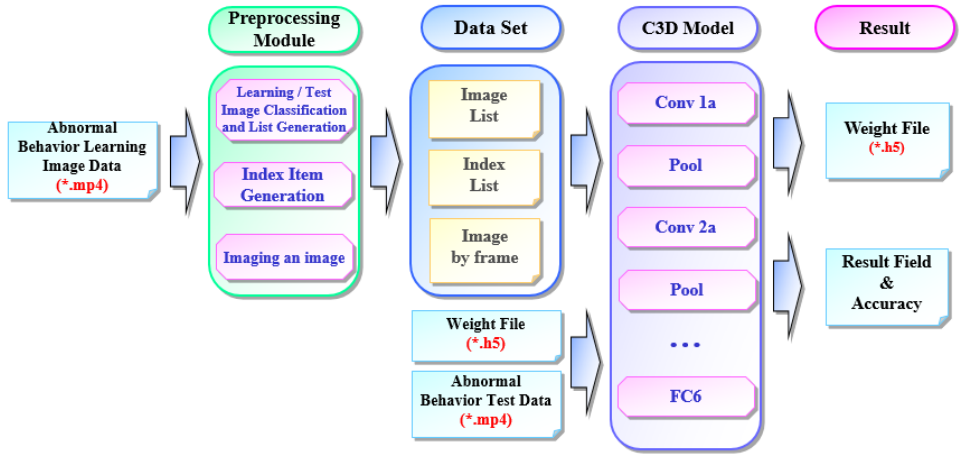

Figure 2. C3D Learning and Detection

3.1 Image

We use the E2ON image dataset for learning and testing. E2ON image data currently consists of five criminal types and 32 detailed types of precursor symptoms and abnormal behaviour. Learning is conducted by dividing into crime types and detailed types.

3.2 Preprocessing Module

When learning a C3D model, various processes are required to apply it to the model. Therefore, the pre-processing process was conducted to establish a data set before learning.

3.2.1 Learning / Test Image Classification and List Generation

For the experiment, a learning image for learning and a test image for a detection test based on the learned result are required. In the classification method, random numbers from 1 to 10 are assigned while reading the video, and if the number of images is greater than 2, it is classified as a test image, and the input image is classified as a learning image and a test image at a ratio of 8 to 2.

Figures 3 and 4 show a list of learning/test images. First, after classifying images, a list of learning / test images to be used to read images in learning should be prepared. In order to prepare a list, the learning/test images were read one by one to check the total frame of the image, and the list of learning/test images was prepared by additionally writing frames in units of 16 frames.

3.2.2 Index Item Generation

There are five crime types and 32 detailed types in the video.

Figures 5 and 6 show crime and detailed type indexes. As shown in Figures 5 and 6, each type is set as an index item and numbered sequentially from 0.

```
sub_train_img/A01/C011_A01_SY01_P01_B01_02DAS_A01 1 0
sub_train_img/A01/C011_A01_SY01_P01_B01_02DAS_A01 17 0
sub_train_img/A01/C011_A01_SY01_P01_B01_02DAS_A01 33 0
sub_train_img/A01/C011_A01_SY01_P01_B01_02DAS_A01 49 0
sub_train_img/A01/C011_A01_SY01_P01_B01_02DAS_A01 65 0
```

```
main_train_img/C011/C011_A01_SY01_P01_B01_02DAS 1 0
main_train_img/C011/C011_A01_SY01_P01_B01_02DAS 17 0
main_train_img/C011/C011_A01_SY01_P01_B01_02DAS 33 0
main_train_img/C011/C011_A01_SY01_P01_B01_02DAS 49 0
main_train_img/C011/C011_A01_SY01_P01_B01_02DAS 65 0
```

Figure 3. Learning Image List

```
sub_test_img/A01/C011_A01_SY01_P01_B01_01DAS_A01 1 0
sub_test_img/A01/C011_A01_SY01_P01_B01_01DAS_A01 17 0
sub_test_img/A01/C011_A01_SY01_P01_B01_01DAS_A01 33 0
sub_test_img/A01/C011_A01_SY01_P01_B01_01DAS_A01 49 0
sub_test_img/A01/C011_A01_SY01_P01_B01_01DAS_A01 65 0
```

```
main_test_img/C011/C011_A01_SY01_P01_B01_01DAS 1 0
main_test_img/C011/C011_A01_SY01_P01_B01_01DAS 17 0
main_test_img/C011/C011_A01_SY01_P01_B01_01DAS 33 0
main_test_img/C011/C011_A01_SY01_P01_B01_01DAS 49 0
main_test_img/C011/C011_A01_SY01_P01_B01_01DAS 65 0
```

Figure 4. Test Image List

3.2.3 Index Item Generation

The assigned number is paired with the image to learn according to the type of the image.

```
0: ChildAbuse(PhysicalAbuse)
1: HomeInvasion(InfrontDoor)
2: Theft(InfrontDoor)
3: Theft(Park)
```

Figure 5. Crime Type Index

3.2.4 Imaging an image

The C3D model proceeds with learning by extracting image features for each frame of the image rather than learning in the form of an image. Therefore, while reading the image, the image is stored in a folder with the image name for each frame, and the data set is constructed by imaging it.

Figure 7 is an example of imaging an image

```
0: A01: Child is alone.
1: A05: Stroller left unattended.
2: A06: Push the stroller hard.
3: A07: Hold the stroller with your feet.
4: A08: Pulled the stroller down hard.
5: A14: Adult throws a child.
6: A16: Overturn the stroller.
7: A17: Pacing around.
8: A18: Trying to open the door lock.
9: A19: kick the door.
10: A20: Trying to look inside the door.
11: A21: Knocked on the door.
12: A29: Hitting with tools.
13: A30: Stealing packages.
14: A31: Hanging around in front of a car.
15: N0: Normal behavior.
16: N1: Normal behavior.
17: SY13: A kid is in a stroller.
18: SY14: A child is walking around.
19: SY15: A person is pacing around the door.
20: SY16: A person is standing around the door.
21: SY17: A person is sitting around the door.
22: SY25: A person is pacing in front of the package.
23: SY26: A person is standing around the delivery.
24: SY28: A person standing in front of a car.
25: SY29: A person is standing around a car.
26: SY30: A person is sitting around a car.
27: SY31: A person is leaning against the door.
28: SY32: A person is leaning against a wall (or pole).
```

Figure 6. Detailed Type Index

Figure 7. Imaging an Image

3.3 Abnormal Behavior Learning

Learning of the C3D model is learned as a frame-specific image of the image. Each of the images is read as shown in the learning and test list. The read images are divided into 16 frames, the features of the images are extracted, and learned in the C3D model with the index and index number generate. The C3D model is mainly learned based on convolution and pooling tasks.

Epoch is the total number of learning using the entire data, and in this paper, up to epoch 10 was allocated and learning was conducted. Figures 8 and 9 show the accuracy and loss rate of the main and detailed types after learning with the allocated epoch. The loss rate of crime types decreased sharply when it reached epoch 1 to 2, and the accuracy started at about 0.96, and rose to about 0.99.

Like the crime type, the loss rate of the detailed type decreased sharply when it became epoch 2 in epoch 1, but it started with a higher loss rate than the crime type and showed a much higher loss rate in epoch 9.

	epoch	loss	acc	val_loss	val_ascc
2	1	1.1564520827240496	0.9589148592213391	0.46716389669294556	0.9392652992602556
3	2	0.3281057572402547	0.9766418563922942	0.30269905672603004	0.9842594149293881
4	3	0.3118234604958429	0.978357806816651	0.2782397119587672	0.9911524882313383
5	4	0.30312080000885155	0.9793041896807221	0.24906924945995215	0.9897024209818427
6	5	0.1805462523289130	0.9944884143877138	0.1617773582841938	0.9953836023313158
7	6	0.16116309978080923	0.9942375050518658	0.14642203220088396	0.9981576440260032
8	7	0.1565182116620868	0.9942543446046073	0.15101892817945017	0.9963363035193903
9	8	0.1537406754765921	0.9944311599083928	0.14012966636210059	0.9978634274826272
10	9	0.13950060757343039	0.9971204364812071	0.13110346050055752	0.9985849585294777
11	10	0.13592989886046322	0.9970210831200323	0.130085097571989	0.9980665770006725

epoch	loss	acc	val_loss	val_acc
1	3.013124878511819	0.9699628507185453	0.15813755463750134	0.9961389961389961
2	0.18992106415173304	0.9871749437872714	0.16667091487545907	0.9947839322839322
3	0.1863556009331154	0.9880731254277055	0.15086956336797788	0.9977168102168102
4	0.17857284595544695	0.9887818946133542	0.14805957687371982	0.9956006831006831
5	0.12089044244912424	0.9972565744452048	0.10978611597400198	0.9968814968814969
6	0.10503056637818238	0.9969205200899404	0.09964980409854875	0.9976239976239977
7	0.09993717679676509	0.997024391436113	0.10122166116831739	0.9969186219186219
8	0.09795791221592641	0.9969082999315672	0.0918729066569939	0.9976239976239977
9	0.09197229530375681	0.9980997653729592	0.09230852748436356	0.9976054351054351
10	0.09062265013436999	0.9982158568775051	0.09065770704818119	0.9977724977724978

Figure 8. Crime Type Accuracy and Loss Rate

epoch	loss	acc	val_loss	val_acc
1	4.385539925671621	0.3299477351916376	2.2617553980952967	0.49554627925322114
2	2.119280060018812	0.5924651567944251	2.014906265689861	0.6515908493294768
3	2.00934091565501	0.6571646341463414	2.010300805529199	0.6746976071522482
4	1.9707686300394012	0.6757273519163763	2.008549671466245	0.6344333420983435
5	1.3771486795737768	0.8057099303135888	1.3238206155796035	0.789409676571128
6	1.2002512595711685	0.8224303135888502	1.286497832284988	0.7885386536944518
7	1.1618904656607931	0.8301829268292683	1.2634847253647221	0.7998783854851433
8	1.1478645933712817	0.833619337979094	1.2686135139146855	0.7995168288193532
9	1.0214888153591222	0.8734581881533101	1.172608056242078	0.8252530896660532
10	0.9846593752306098	0.8790026132404182	1.158826397761777	0.8246778858795688

	epoch	loss	acc	val_loss	val_acc
2	1	2.6717527373356753	0.5230642637283051	1.5863389641056467	0.6666296008006227
3	2	1.6454005408732915	0.6533942474328134	1.655992658579958	0.6512078839097075
4	3	1.6279999833658192	0.6596424691549624	1.5297365464318995	0.7020112865562104
5	4	1.6198276917904384	0.6615710923152531	1.5307349442282416	0.6912181696875348
6	5	1.193077501997736	0.7360566850314271	1.125877192384155	0.754614700322473
7	6	1.1305491841348771	0.751715229818163	1.1148114285598432	0.7618147448015122
8	7	1.1201205833505656	0.7600973849615892	1.1055209319609816	0.7722186700767263
9	8	1.1157498674091941	0.7655874136726935	1.133190346595298	0.761321305459802
10	9	1.0012703748478997	0.8024543467577145	1.0011094027448122	0.8019987768264205
11	10	0.968205141873482	0.8095383590698637	1.0158953472423458	0.7929987212276215

Figure 9. Detailed Type Accuracy and Loss Rate

The accuracy of the detailed type also began with a much lower accuracy than the crime type, and showed a significant change as the epoch was repeated, but it was about 0.8 and showed a much lower accuracy than the crime type. It is judged that the detailed types showed lower accuracy and higher loss rate due to similar types at first glance, such as hanging around the door and standing in front of the door, with more index items of the type than the crime type.

3.4 Abnormal Behavior Detection

The detection of abnormal behavior using the C3D model proceeds with an image obtained by dividing the test image into 16 frame pieces and a weight file with weights stored by learning using the C3D model. The 16-frame images are tested by scoring and detecting which type is closest.

Figure 10 shows an example of the type detection test result for each frame after abnormal behavior detection using the C3D model. It shows the type of crime detected from above, the accuracy of the type, and the type of crime that should have been detected. Subsequently, the detailed type and the accuracy of the type are presented, and in the case of detailed items, the detailed items are displayed exactly what behavior it is. And the results of ranking 1 to 5 in order of accuracy are shown once more. Finally, the original detailed type appears.

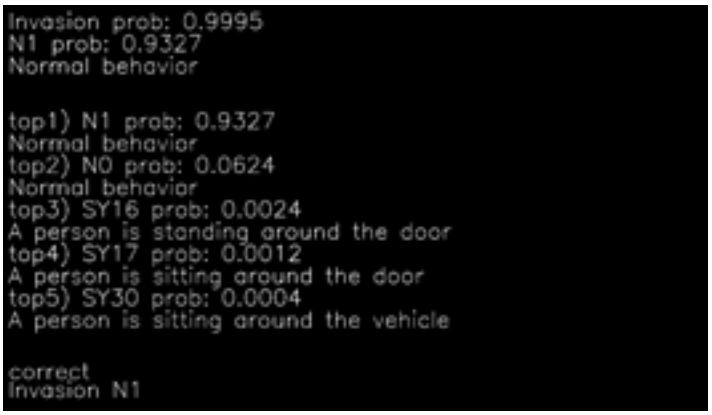

Figure 10. Example of Type Detection Test Results by Frame

4. Experimental Results and Analysis

Experiments in this paper were conducted using geforce rtx 2080 with approximately 11 gigabytes of memory and python in linux environments. In the case of abnormal behavior classification, the test is conducted by dividing the test image into 16 frame pieces, scoring which type is closest, and then classifying. Figure 11 shows the results of the abnormal behavior test by crime type.

Figure 11 shows the test results of the theft (parking lot) image.

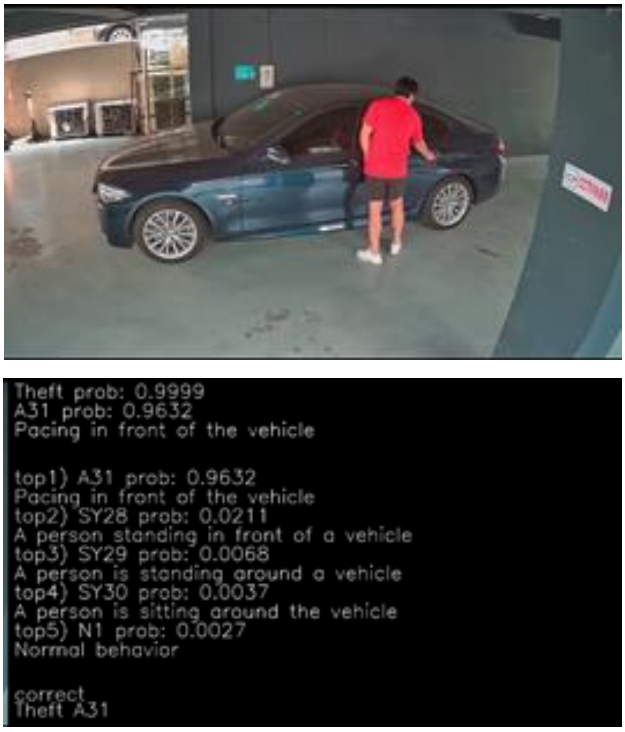

Figure 11. Theft (Parking Park) Test Results

The results were visible in all frames of the image, and one of them was examined. The crime type was detected according to theft (parking lot), and the accuracy was about 99%. As for the detailed type, it could be seen that the current A31 was detected with about 96% accuracy when he was hovering in front of the vehicle.

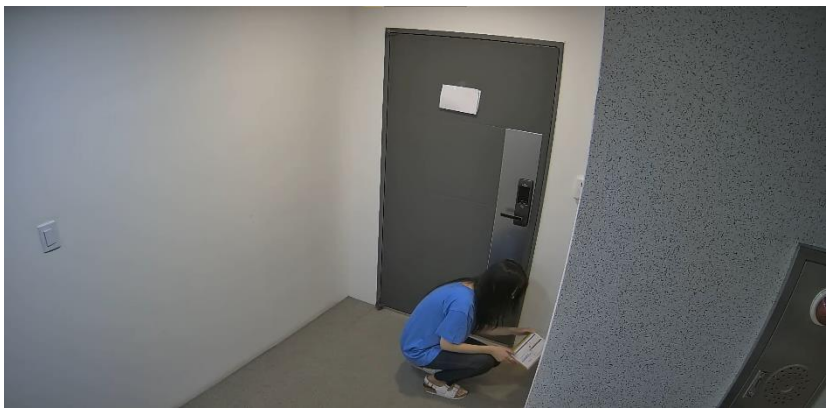

Figure 12. Data Source Image for Learning

Figure 13 shows the test results of the intrusion (in front of the door) image. The results were visible in all frames of the image, and one of them was examined. The crime type was detected according to theft (in front of the door), and the accuracy was about 99%. As for the detailed type, it was found that the detection was correct with an accuracy of about 94% when the current N1 normal behavior was performed.

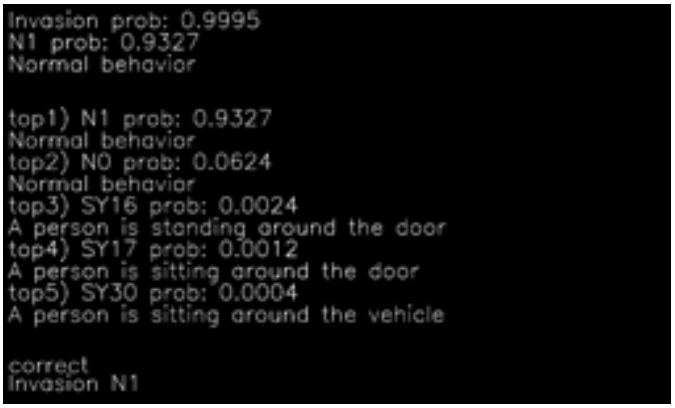

Figure 13. Residential Break-in(in front of Door) Test Results

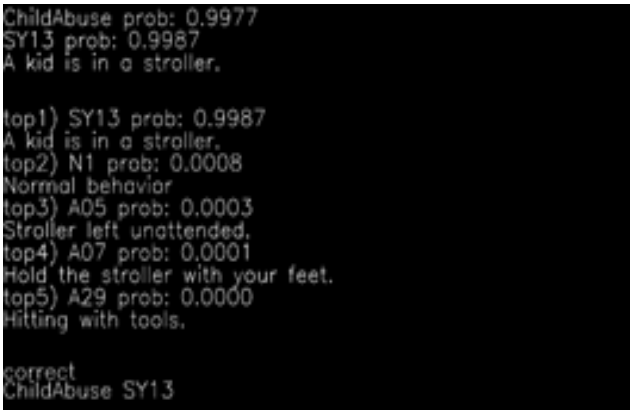

Figure 14. Child Abuse Test Results

Figure 14 shows the test results of child abuse (indulgence) images. The results were visible in all frames of the image, and one of them was examined. The crime type was detected according to child abuse (indulgence), and the accuracy was about 96%. As for the detailed type, it was currently SY01, which was detected with about 22% accuracy when the child was pacing alone.

Figure 15. Theft (in front of the Door) Test Results

Figure 15 shows the test results of the theft (in front of the door) image. The results were visible in all frames of the image, and one of them was examined. The crime type was detected as theft (in front of the door), and the accuracy was about 99%. As a detailed type, it was possible to examine the detection that the A30 was classified with approximately 99% accuracy with prob: 0.9903 when the current A30 courier was stolen.

5. Conclusion and Future Research

Abnormal behavior refers to behavior that is not socially valuable or suitable for daily life. Most of these actions are criminal acts that people fear. To prevent this from happening, a study was conducted to learn and detect abnormal behavior in CCTV images using the C3D model, an artificial intelligence model. Among the learning models for anomaly behavior detection, C3D is a model that trains 3D Convolution Networks on large amounts of image data in a supervised learning manner. For this reason, the C3D model, an artificial intelligence model, was adopted.

As a result of learning and detecting abnormal behavior images in CCTV with the C3D model, which is an artificial intelligence model, abnormal behavior was detected from various crime type image data, showing quite high accuracy. Therefore, it is believed that abnormal behavior detection technology using C3D can be applied to various CCTV image data to detect the precursor symptoms of crime, prevent crimes in advance, or detect criminal acts to protect victims.

In addition, it is believed that a healthy and safe society can be created if the C3D model can be used in the medical community as a method of learning and classifying the precursor symptoms or abnormal behavior of the disease by applying it to medical images.

Acknowledgements

This research was supported by Seokyeong University in 2022.

References

Bae, H., et al., LSTM(Long Short-Term Memory)-Based Abnormal Behavior Recognition Using AlphaPose, KIPS Transaction on Software and Data Engineering, Vol. 10, No. 5, pp. 187-194, 2021. https://doi.org/10.3745/KTSDE.2021.10.5.187

Chalapathy, R., Chawla, S., Deep Learning for Anomaly Detection: A Survey. arXiv preprint arXiv:1901.03407, 2019. https://doi.org/10.48550/arXiv.1901.03407

Chandola, V., et al., Anomaly Detection : A Survey, ACM Computing Surveys, Vol. 41, No. 3, pp. 1-58, 2009. https://doi.org/10.1145/1541880.1541882

Haroon, U., A Novel 3D-Convolution Neural Network for Human Interaction Recognition in Videos, The Journal of KINGComputing, Vol.18, No.1, pp. 19-28, 2022. http://doi.org/10.23019/kingpc.18.1.202202.002

Hong, C., & Choi, S., ROC curve generalization and AUC, Journal of the Korean Data & Information Science Society, Vol. 31, No. 4, pp. 477-488, 2020,

Ji, S., et al., 3D Convolutional Neural Networks for Human Action Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 1, pp.221– 231, 2013. doi: 10.1109/TPAMI.2012.59.

Kiran, B., et al., An Overview of Deep Learning based Methods for Unsupervised and Semi-supervised Anomaly Detection in Videos, Journal of Imaging, Vol. 4, No. 36, pp. 1-25, 2018. doi: 10.3390/jimaging4020036

Lee, J., A Study on the Implementation of Intelligent Abnormal Behavior Monitoring System Using Deep Learning, Hanse Univ., Ph.D Thesis, 2020.

Lee, J., Lee, K., An Anomalous Sequence Detection Method Based on An Extended LSTM Autoencoder, The Journal of Society for e-Business Studies, Vol. 26, No. 1, pp. 127-140, 2021.

Leng, B., et al., 3D object understanding with 3D Convolutional Neural Networks, Information sciences v.366, 2016 년, pp.188 – 201

Malhotra, P., LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection, ICML 2016 Anomaly Detection Workshop, arXiv:1607.00148, 2016. https://doi.org/10.48550/arXiv.1607.00148

Maturana, D., & Scherer, S., VoxNet: A 3D Convolutional Neural Network for Real-time Object Recognition, 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, pp. 922-928, DOI: 10.1109/IROS.2015.7353481.

Morais, R., Learning Regularity in Skeleton Trajectories for Anomaly Detection in Videos, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11996-12004, 2019.

Park, H., A Study on Monitoring System for an Abnormal Behaviors by Object's Tracking, Journal of Digital Contents Society, Vol. 14, No. 4, pp. 589-596, 2013. http://dx.doi.org/10.9728/dcs.2013.14.4.589

Park, S., et al., Anomaly Detection by a Surveillance System through the Combination of C3D and Object-centric Motion Information, Journal of KIISE, Vol. 48, No. 1, pp. 91-99, 2021.

RNN Structure, https://velog.io/@sksmslhy/RNN

Tran, Du., et al., Learning Spatiotemporal Features with 3D Convolutional Networks, Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 4489-4497, 2015.

Wang, P., et al., 3D Shape Segmentation via Shape Fully Convolutional Networks, Computers & Amp; Graphics, Elsevier BV, Vol. 76, pp. 182–192, 2018. https://doi.org/10.1016/j.cag.2018.07.011

Wu, T., & Lee, E., Human Action Recognition Based on 3D Convolutional Neural Network from Hybrid Feature. Kournal of KMMS, Vol. 22, No. 12, pp. 1457-1465. 2019. https://doi.org/10.9717/kmms.2019.22.12.1457

Zeng, H., Learning-Based Multiple Pooling Fusion in Multi-View Convolutional Neural Network for 3D Model Classification and Retrieval, Journal of Information Processing System, Vol. 15, No. 5, pp. 1179~1191, 2019. doi: 10.3745/JIPS.02.0120