https://doi.org/10.21289/KSIC.2023.26.2.217

YOLOv5 based Anomaly Detection for Subway Safety Management Using Dilated Convolution

Nusrat Jahan Tahira[1], Ju-Ryong Park[1], Seung-Jin Lim[2], Jang-Sik Park[3*]

〈Abstract〉

With the rapid advancement of technologies, need for different research fields where this technology can be used is also increasing. One of the most researched topic in computer vision is object detection, which has widely been implemented in various fields which include healthcare, video surveillance and education. The main goal of object detection is to identify and categorize all the objects in a target environment. Specifically, methods of object detection consist of a variety of significant techniques, such as image processing and patterns recognition. Anomaly detection is a part of object detection, anomalies can be found various scenarios for example crowded places such as subway stations. An abnormal event can be assumed as a variation from the conventional scene. Since the abnormal event does not occur frequently, the distribution of normal and abnormal events is thoroughly imbalanced. In terms of public safety, abnormal events should be avoided and therefore immediate action need to be taken. When abnormal events occur in certain places, real time detection is required to prevent and protect the safety of the people. To solve the above problems, we propose a modified YOLOv5 object detection algorithm by implementing dilated convolutional layers which achieved 97% mAP50 compared to other five different models of YOLOv5. In addition to this, we also created a simple mobile application to avail the abnormal event detection on mobile phones.

Keywords : Object Detection, Convolutional Network (CNN), Dilated Convolutional Layer, Anomaly Detection

1 Dept. of Electrical, Electronics and Communications Engineering, Kyungsung University
2 Busan Transportaion Corporation
3* Corresponding Author, Dept. of Electronic Engineering, Kyungsung University, Professor
E-mail: jsipark@ks.ac.kr

1. Introduction

Recently anomaly detection has become a notable but challenging task in computer vision and pattern recognition. With the expansion of technologies, researchers started using the deep learning based methods for object detection and classification in fields such as video surveillance[1, 2], healthcare [3] and public transportation[4]. In this however, normal and anomaly events detection is a task of interest for research lately. Anomaly detection[5] can be described as a binary classification between normal and abnormal events. As abnormal events do not occur frequently, it is very difficult to detect these kinds of events. The human brain can easily distinguish between normal and abnormal objects in an image, but it is not so trivial for a machine.

Now-a-days, deep Convolutional Neural Network[6, 7] based models are obtaining high accuracy in anomaly detection, because CNN can automatically learn features of objects from a large set of labelled data[8]. Among all the deep learning based method, one of the most popular algorithm for real-time object detection is YOLO (You Only Look Once)[9] that achieved promising performance. In our work, we focused on detecting and classifying four classes namely assault, fall-down, normal and vandalism for the task of anomaly detection. Our main contribution in this research is the use of dilated convolutional layers [10] in YOLOv5 which brings about an improvement in anomaly detection performance compared to the five different YOLOv5 models.

The following chapters describes the dilated convolution and anomaly detection model architecture. We further elaborate the implementation of the mobile application for the anomaly detection task and results from the experiment, and conclude with what we hope to accomplish in the future.

2. Methodology

2.1 Dilated Convolutional Layer

Convolution is a mathematical operation that allows combining of two sets of information. This is applied to the model input data to filter the information and produce a feature map, and the filter can be called a kernel. To perform convolution, the dimension of kernel can be, for example 3×3, which goes over the input image and perform the matrix multiplication element after element [11].

By inserting holes between the kernel's consecutive elements, dilated convolution can expand the kernel size. We can simply say that, it is the same as a standard convolution but consists of skipping layers [12]. Mathematically,

$$(F * k)(p) = \sum_{s+t=p} F(s)k(t) \qquad (1)$$

$$(F *_l k)(p) = \sum_{s+lt=p} F(s)\,k(t) \qquad (2)$$

where

$F(s) = \in put,$

$k(t) = Applied\,Filter,$

$*_l = l - dilated\,convolution,$

$(F *_l k)(p) = Output$

Equation (1) is the standard convolution while (2) represents dilated convolution. As seen in equation (2), the summation, $s+lt=p$ indicates that we will skip some points during convolution as the parameter l is known as the dilation rate which tells us how much we want to widen the kernel. As it skips the pooling steps, it consumes less memory while implementing the algorithm and covers a much more wider area than the basic convolution.

When $l=1$, it is the standard convolution but it will be dilated when $l>1$. In our implementation, we used a 3×3 kernel filter and dilation rate of 2, so it has the same view as 5×5 kernel filter while using 9 parameters as seen in Fig. 1.

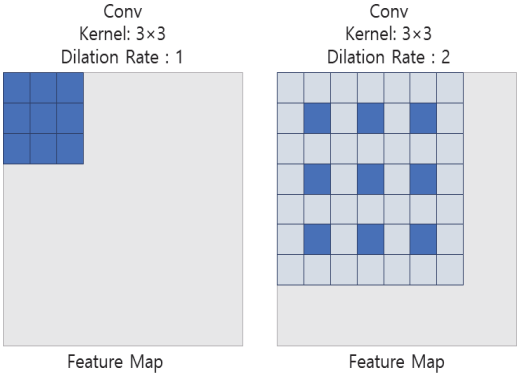
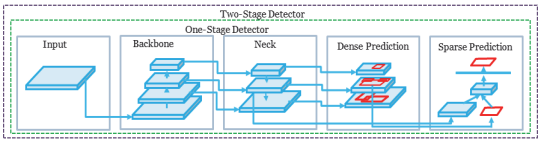


Fig. 1 Dilated convolution

2.2 Feature Extraction

The process of transforming raw data into numerical features that can be processed while preserving the information in the original dataset is described as feature extraction [13]. Feature extraction helps to reduce the amount of unnecessary data from the dataset which is useful to increase the speed of learning and generalization steps in the model learning process and also to build the model with less computation cost.

From Fig. 2, we can see that there are three main parts in the object detection framework; backbone, neck and head. For extracting key features from the given input, model backbone is mainly used. In our model, we used CSP-DarkNet53 [14] as the backbone to extract rich informative features. In dealing with time and deeper networks, CSP-DarkNet53 has shown an noticeable improvement[14]. Necessarily, a model neck is utilized to get feature pyramids. For identifying the same object or person with different sizes and scales, features pyramids plays a great role and we used PANet [15] in



Input: {Image, Video, Patches, Image Pyramid,...}
Backbone: {VGG16, ResNet-50, DarkNet53,...}
Neck: {FPN, PANet, Bi-FPN,...}
Head:
Dense Prediction: {YOLO, SSD, RPN,...}
Sparse Prediction : {Faster R-CNN, R-FCN,...}

Fig. 2 Object detection framework

this implementation.

Model head performs the final detection. It outputs the final vectors with class probabilities, objectness scores and bounding boxes by applying anchor boxes on features. In our implementation we used YOLO layer as model head.

The proposed model architecture based on YOLOv5 with dilated convolutional layer. The model parameters used in our implementation include the Leaky-ReLU activation function [16] and Adam optimizer. Besides, Binary Cross-Entropy [17] with Logits Loss function from Pytorch are used to calculate the class probability and object score.

3. Implementation and Results

Our anomaly detection dataset includes ground truth for four different activities (assault, fall-down, normal and vandalism) shown in Fig. 4. The data was collected from AI Hub and we created the ground truth (GT) using Roboflow. The data was split into 73:17:10 corresponding to training, validation and testing respectively. Table 1 represents the model layers and parameters of the proposed model used in this experiment which is same as the five different YOLOv5 models.

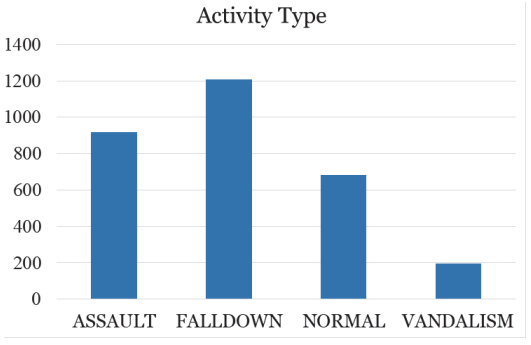


Fig. 4 Dataset distribution

Table 1. Model layers and parameters

Models	Layers	Parameters (M)
YOLOv5n_dilated(ours)	157	1.9
YOLOv5s_dilated(ours)	166	7.2
YOLOv5m_dilated(ours)	212	21.2
YOLOv5l_dilated(ours)	267	46.5
YOLOv5x_dilated(ours)	322	86.7

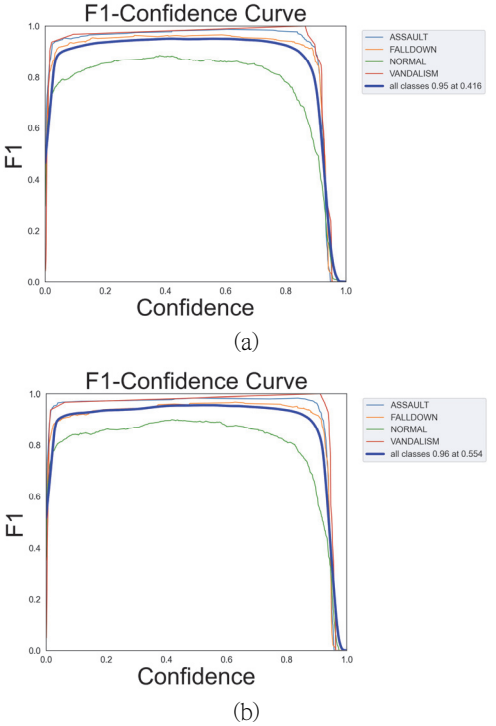


Fig. 5 Comparison of F1-confidence curver
(a) F1-Confidence Curve of YOLOv5n without dilation
(b) F1-Confidence Curve of YOLOv5s with dilation

Fig. 5(a) shows the F1-Confidence Score of YOLOv5n model without dilation whereas Fig. 5(b) shows F1-Confidence Score of the same model with dilated convolutional layer. we can see that model without dilation achieved 95% performance at 0.416 IoU threshold. On the other hand, the model with dilation achieved 96% performance at 0.554 IoU. We can therefore say that, the model performs better with dilated convolution layers.

Fig. 6(a) shows confusion matrix of YOLOv5n model without dilation whereas Fig. 6(b) shows confusion matrix with dilated convolutional layers.

In the above confusion matrices, background images are images with no objects that are added to the dataset to reduce False Positives (FP). No labels are required for this. The equations that are used to calculate the model performance are given below,

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$mAP = 1/n^* \sum(AP)$$

Where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. The mAP is mean average precision where n is number of classes and AP is average precision.

Fig. 7 shows the mAP50 of all five YOLOv5 models with our proposed model. we can see that proposed model with dilation is better than the models without dilation except YOLOv5s.

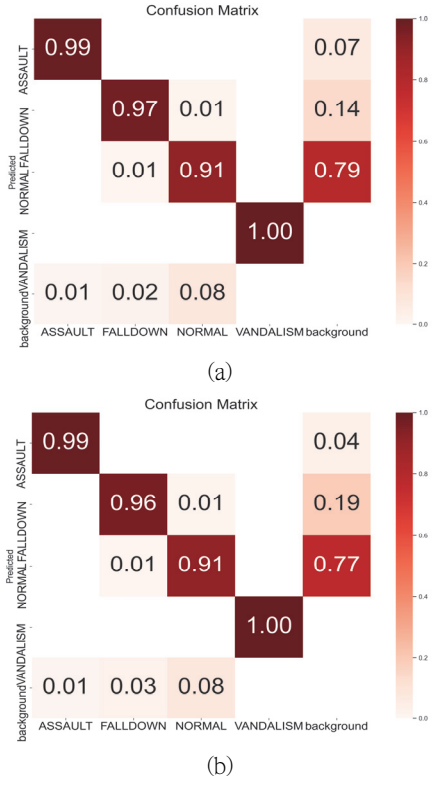


Fig. 6 Comparison of confusion matrix
(a) Confusion matrix without dilation (b) Confusion matrix with dilation

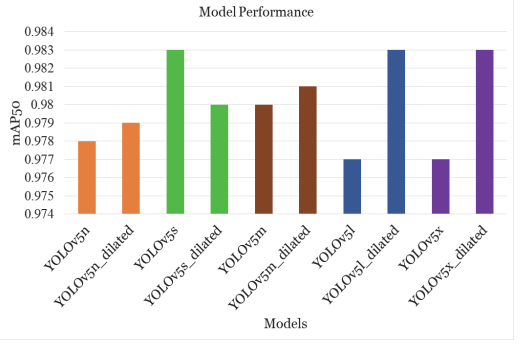


Fig. 7 Model performance

4. Conclusions

Object detection is a leading skill for most computer systems. Although there are many object detection techniques seen in the last few years, some models have achieved outstanding performance. Anomaly detection is also a part of object detection where we can classify abnormal or normal events in an image or video. As abnormal events do not occur frequently, it is hard to detect abnormal events. In this work, we implemented the five different YOLOv5 models and proposed a YOLOv5 based model with dilated convolution to detect anomaly events in the metro which achieved 97% mAP50. This model is very fast and easy to use not only for study purposes but also in real time.

In future, we plan on testing and implementing the system in real time scenarios using the CCTV surveillance cameras.

Acknowledgements

This work was supported by the BB21+ funded by Busan Metropolitan City and Busan Institute for Talent & Lifelong Education(BIT) and supported by "Human Activity Data of Unmaned Store" of AI learning data construction project through NIA(National Information Society Agency)

References

[1] M Paul, M. E. Shah and S. Chakraborty, "Human detection in surveillance videos and its applications-a review", EURASIP Journal on Advances in Signal Processing, pp. 176, 2013.

[2] Lo, Win-Tsung, Yue Shan Chang, Ruey-Kai Sheu, Chao-Tung Yang, Tong-Ying Juang, and Yu-Sheng Wu. "Implementation and evaluation of large-scale video surveillance system based on P2P architecture and cloud computing." International Journal of Distributed Sensor Networks 10, no. 4, 2014

[3] Litjens, G., et al.: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci. Reports 6 (2016)

[4] Thomas Defard, Aleksandr Setkov, Angelique Loesch, Romaric Audigier, "PaDiM: a Patch Distribution Modeling Framework for Anomaly Detection and Localization", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2020.

[5] F. Mercaldo, F. Martinelli, and A. Santone, "A proposal to ensure social distancing with deep learning-based object detection," in 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021, pp. 1–5.

[6] Krizhevsky A., Sutskever I., Hinton G.E., Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems (2012), pp. 1097-1105.

[7] Zhang Q., Yang L.T., Chen Z., Li P. , A survey on deep learning for big data, Inf. Fusion, 42 (2018), pp. 146-157.

[8] Yuan J., Hou X., Xiao Y., Cao D., Guan W., Nie L., Multi-criteria active deep learning for image classification, Knowl.-Based Syst., 172 (2019), pp. 86-94

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once:

Unified, Real-Time Object Detection., published at Computer Vision and Patter Recognition (CVPR)., 2015.

[10] Fisher Yu, Vladlen Koltun. “Multi-Scale Context Aggregation by Dilated Convolutions” Computer Vision and Pattern Recognition; 2016.

[11] Yulia Gavrilova, 'Convolutional Neural Networks for Beginners', Serokell, August 3rd, 2021, https://serokell.io/blog/introduction-to-convolutional- neural-networks

[12] 'Dilated Convolution', Geeksforgeeks, 02 Mar, 2022, https://www.geeksforgeeks.org/dilated-convolution/

[13] Benyamin Ghojogh, Maria N. Samad, Sayema Asif Mashhadi, Tania Kapoor, Wahab Ali, Fakhri Karray, Mark Crowley, “Feature Selection and Feature Extraction in Pattern Analysis: A Literature Review”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[14] Alexy Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao; “YOLOv4: Optimal Speed and Accuracy of object Detection”. Computer Vision and Pattern Recognition(CVPR); 23 April 2020.

[15] Liu S.,. Qi L., Qin H., Shi J., Jia J. Path aggregation network for instance segmentation; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Salt Lake City, UT, USA. 18-23 June 2018; pp. 8759-8768.

[16] Shiv Ram Dubey, Satish Kumar Singh, Bidyut Baran Chaudhuri, “Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2021.

[17] Zhilu Zhang, Mert R. Sabuncu, “Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018.

(Manuscript received December 7, 2022;
revised December 27, 2022; accepted December 30, 2022)