Design and Implementation of Real-time Anomaly Detection System based on YOLOv4

DOOHWAN KIM[1,2], YO-HAN HAN[1,2], JONGPIL JEONG[1,*]

[1]Department of Smart Factory Convergence, Sungkyunkwan University
2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do, 16419,
REPUBLIC OF KOREA

[2]AI Machine Vision Smart Factory Lab, Dev
296, Sandan-ro, Danwon-gu, Ansan-si, Gyeonggi-do, 15433,
REPUBLIC OF KOREA

Abstract: - To solve the problem of high-wage employment and unemployment that is constantly occurring in industrial sites, we designed a real-time anomaly detection system based on YOLOv4 to automate the detection of defective products at actual manufacturing sites. This contributes to reducing labor costs and increasing work efficiency in the field. It also contributes to manufacturing data collection and smart factory system construction by utilizing the established system

Key-Words: - AI Deep Learning, Anomaly Detection, Smart Factory, Supervised Learning, YOLOv4, Edge Computing, Manufacturing Data Platform

Received: July 22, 2021. Revised: October 11, 2022. Accepted: November 14, 2022. Published: December 29, 2022.

1 Introduction

With the advent of Industry 4.0, new changes are coming to the manufacturing industry as IT technology develops. The application of smart factories based on innovative technologies for manufacturing competitiveness cannot be delayed any longer, [1], [2]. Smart factories are meaningful in that they meet various customer requirements and improve production efficiency by collecting and analyzing production information in real-time from the manufacturing site by breaking away from the existing mass production method.

After coronavirus Disease-2019 (COVID-19), the preference for non-face-to-face work and the rise in labor costs have made it difficult to secure manufacturing manpower at the manufacturing site. Recently, to solve this problem, there is a demand to entrust the product quality inspection task to an AI-based vision inspection system, [3]. As it is necessary to perform the task of determining whether the product is defective or not Must have a high level of accuracy Since the appropriate cost has to be calculated at a level that can be introduced even in small and medium-sized enterprises, a high-quality, low-cost and efficient system is required.

This study is YOLOv4, [4], based on the anomaly detection, [5], inspection method, [6], it inspects objects moving on the conveyor belt in real-time and aims to determine whether the object is good or bad, and to classify products. The proposed system meets the level of product defect inspection required by actual manufacturing sites. It can be expected to have the effect of improving productivity and the operating cost of the company, [7].

In addition, it aims to link the system to the Manufacturing Execution System (MES) and Enterprise Resource Planning (ERP) systems used in the factory by building the system on an edge computing device.

The established system supports the monitoring system to adjust factory production schedules or respond to emergencies using the collected data while collecting manufacturing data. Through this, it contributes to building a smart factory that can be controlled remotely, and a deep learning trained model can be applied in real-time in a cloud environment. And uploading the collected data contributes to building a manufacturing data platform so that manufacturing data can be used more innovatively, [8].

The structure of this paper is as follows. Section 2 describes the related research, and Section 3 describes the design of the real-time anomaly detection system applied in this paper. Section 4 shows the experimental results of this system, and Section 5 concludes with a conclusion.

2 Related Work

This section outlines the concepts or techniques used in papers such as anomaly detection or YOLOv4

2.1 Anomaly Detection

Anomaly detection refers to the problem of distinguishing between normal samples and abnormal samples. technology being applied. When both normal and abnormal samples exist by labeling in a given training data set, it is called Supervised Learning, [9].

When a model trained in this way identifies an object, it is called supervised anomaly detection. This way high It is mainly used when accuracy is required, The more diverse the normal and abnormal samples are, the higher the performance can be derived, [10].

In the case of semi-supervised learning, it means that only normal classes are defined and learned. This can be useful when you want to detect a class other than the normal class. In the case of unsupervised learning, no class is defined and only images are learned to induce the deep learning model to make its judgment. As a separate class is not defined, the learning image aims to use only the normal class or train only the bad class. However, it is recommended to train only one class of images to be learned, [11].

In this paper, supervised learning is applied to the
above concepts. There are many cases where only one type of defect class is not defined, and each defect type has a different method to be taken on-site. For example, if a product is not cut correctly, it may be necessary to correct the cut. Likewise, if the color of the product has changed, the material of the product may require checking the injection temperature.In addition, statistics on what kind of defects should occur a lot should be available, and since high accuracy is required, supervised learning was selected over semi-supervised or unsupervised learning even if there were difficulties in learning.

2.2 YOLOv4

You only look once (YOLO) is a state-of-the-art, realtime object detection system. R-CNN, [12], which is representative object detection, generates and learns about 2,000 Region Proposals when learning images. This consumes a lot of time and resources for learning speed. To compensate for these shortcomings, the YOLOv4 model appeared, and real-time, Highquality and reliable object detection results can be obtained. Fig. 1 shows the labeling of predefined objects in images using the YOLOv4 Model.

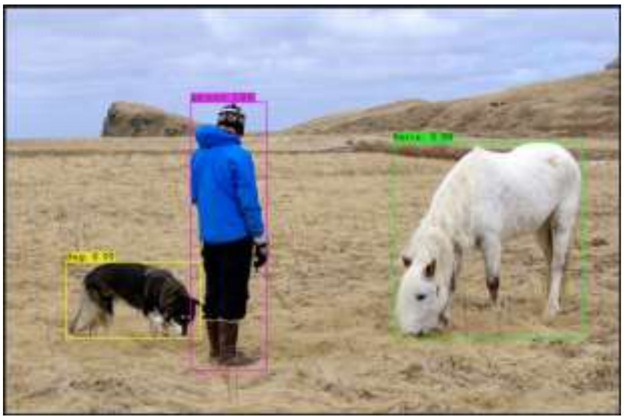

Fig. 1: Detected Object by YOLOv4

The image acquired on the conveyor belt must be discriminated in real-time, and the discriminating result must be delivered to the discharge unit. In addition, a long learning rate may interfere with the input of the field that changes flexibly. Therefore, in this paper, the YOLOv4 model with fast learning speed and discrimination speed was adopted.

2.3 Edge Computing

Edge computing, [13] refers to performing computing at or near the physical location of a user or data source. By processing computing services at a location close to the user's end device, users will receive faster and more reliable services, and businesses will benefit from flexible hybrid cloud computing. Edge computing is one way for enterprises to distribute data computation and processing in multiple locations using a common pool of resources, [14].

Image learning can be performed in the cloud or on a remote server. However, if image identification is performed through a remote server or a cloud server, product identification and communication with the programmable logic controller (PLC) may not be performed smoothly due to network speed and various variables. Accordingly, in this paper, an edge computing device was built next to a conveyor belt where products are produced. In addition, it may be difficult to prepare a separate learning server in a small-scale factory. Accordingly, a device capable of learning images was prepared. Additionally, the Monitoring User Interface (UI) Program was prepared to be executed so that the user can check the determined image in real-time. The device is connected to the Internet and is ready to be linked to the factory's MES, ERP system at any time. In addition, a system that can remotely check for product defects at the site and instruct production, suspension, and change of products in real-time has been linked.

DOI: 10.37394/232017.2022.13.17

2.4 Precision-Recall

The Confusion Matrix is the basis for determining how accurate the learning of the AI model is, [15], [16]. To understand the Confusion Matrix, you need to understand 4 concepts first. True Positive (TP) : Predicts that an answer that is actually True is True (correct answer), False Positive (FP) : Actual False Predict that the correct answer is True (incorrect answer), False Negative (FN) : Predict the correct answer that is True as False (Answer incorrect), True Negative (TN) : Predict the correct answer that is actually False as False (correct answer) It shows intuitively, [16]. Recall or Sensitivity (as it is called in Psychology) is the proportion of Real Positive cases that are correctly Predicted Positive. This measures the Coverage of the Real Positive cases by the +P (Predicted Positive) rule. Its desirable feature is that it reflects how many of the relevant cases the +P rule picks up. Conversely, Precision or Confidence (as it is called in Data Mining) denotes the proportion of Predicted Positive cases that are correctly Real Positives. This is what Machine Learning, Data Mining and Information Retrieval focus on, but it is totally ignored in ROC analysis. It can however analogously be called True Positive Accuracy (TPA), being a measure of the accuracy of Predicted Positives in contrast with the rate of discovery of Real Positives, [17].

In this paper, we will apply the concepts of precision and Recall to determine whether products identified by deep learning training have been accurately classified. For accurate learning and classification of deep learning models, the objective model evaluation must be performed by grafting the corresponding concept.

2.5 Manufacturing Data platform

While global manufacturing is becoming more competitive due to variety of customer demand, increase in production cost and uncertainty in resource availability, the future ability of manufacturing industries depends upon the implementation of Smart Factory. With the convergence of new information and communication technology, Smart Factory enables manufacturers to respond quickly to customer demand and minimize resource usage while maximizing productivity performance, [8].

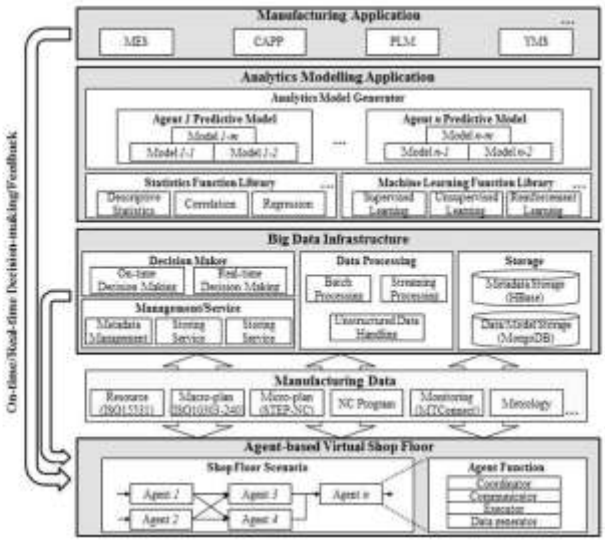

Fig. 2: A big data analytics platform architecture for manufacturing systems, [8].

In addition, there are also cases where new smart factories have been established based on manufacturing data in introducing companies that focus on the personalized cosmetics manufacturing industry, [18]. Manufacturing data is now attracting attention as useful data that cannot be discarded in the 4th industrial era, and it is important to secure data collected in all production processes. Fig. 2 shows the smart factory platform architecture of the big data-based manufacturing site system.

It collects good and bad production of products through edge computing equipment. It also collects product production information and set information. It can be used to build a manufacturing data platform using the collected information and apply it to build an artificial intelligence-based smart factory system.

3 YOLOv4-Based Real-Time Anomaly Detection System

In this paper, the composition of an optical system that can determine whether a sticker attached to the bottom of a cosmetic container is defective among plastic injection (molded) products and image reading through deep learning training was studied.

Shows the conceptual diagram of the system proposed in this paper. Due to the material characteristics of the inspected product Since diffuse reflection occurs due to the reflection of light, it is configured as an indirect lighting type that minimizes reflection of light for optimal configuration of the optical system. Similarly, to avoid diffuse reflection of the product, the camera angle was corrected by 2 to 5 degrees so that it was not perpendicular to the product, [19].

When the proposed system is run through an edge computer, the installed camera continuously acquires images with softwareGrab, [20], in the acquired area (FOV, [21]) Check that the product is detected using a Hough Circle Transform, [22]. When a product is detected, the coordinates of the photographed product are calculated to check whether the product passes through a discernable location. If the product passes through a discriminable location, the learned YOLOv4 Model determines whether the product is defective. The determined image and result are output from the separately implemented monitoring system The determination result is transmitted to the discharge cylinder. Since the discharge cylinder only moves the cylinder according to the electrical signal of the program, After storing the determination result in the edge computer's determination program, when the determined product passes in front of the cylinder, the product determination result is transmitted. As a result of the transfer, the defective product is pushed through the cylinder to be moved to a different conveyor belt line than the normal product. To prevent further damage to other normal or defective products in the process of pushing the product, a sponge is attached to the cylinder to alleviate the impact when the cylinder and product collide. Fig. 3 shows the concept described above as a figure.

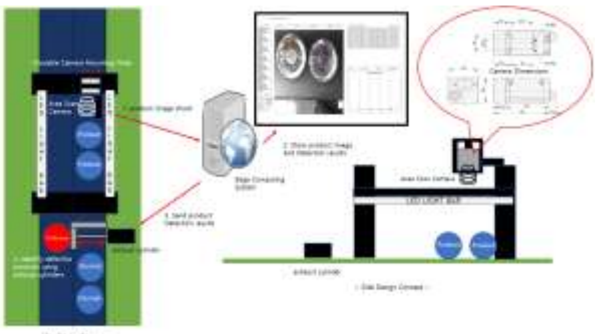

Fig. 3: System Design Concept.

4 Experimental Results

4.1 Experiment Environment

Table 1 describes the list of Hardware used to design the system. An acA1300-60gm mono camera sold by Balser was used as the camera, and a 16mm C-mount lens sold by Tamron was used as the camera lens. The monitoring system and the edge computer that performs deep learning and image discrimination were prepared with the same specifications as above. For optical system lighting, two LED bar lights were used.

Table 1. Real-Time Anomaly Detection System Hardware Spec.

LENS	Manufacturer Part Number	Tamron M118FM16
	Focal Length	16mm
	Lens Mount	C-mount
	Optical Format	1/1.8”
CAMERA	NAME	acA1300-60gm
	Sensor Format	1/1.8”
	Sensor Type	CMOS
	Mono or Color	Mono
IPC	Processor	AMD RyZen 5 5600X 6-Core Processor 3.70 GHz
	RAM	32 GB
	GPU	NVIDIA GeForce RTX 2080 Ti
Optical Lighting	Product /Model Name	LED T5/HF-T5030
	Rated voltage	220V 60Hz 5W
	Spec	300 x 22.5 x 33 (mm)

The test site was conducted on a conveyor belt where actual products are produced. The specifications and test environment of the products used are introduced in Table 2, The learned labeling class shown in Fig. 5 can be checked.

As a special feature, since we plan to read two products in one image, we defined the distance between the products and set the working distance between the product and the camera a little further than when shooting one product. In addition, if the product moves too fast, it is difficult to acquire images, so the movement speed is specified. However, if the moving speed needs to be increased for the production speed of the product, it can help to acquire accurate images by adjusting the settings of the camera and lens.

Table 2. Cosmetic Container and Spec Experimental Environment Setting Value.

Diameter	73.8mm
Height	44.4mm
Sticker diameter	25.2mm
Distance between products	6~10mm
Working Distance	450mm
Conveyor Belt Speed	50mm/s

As explained in Section 3, since the inspection material of the product is a sticker, reflection occurs when light shines directly on the product, so the image is not accurately acquired. Therefore, by correcting the angle of incidence of the light illuminating the product, the product is induced to be illuminated with indirect light. In addition, when the camera shoots a product vertically, light reflection may also occur, so adjust the angle of the camera so that light reflection does not occur. Fig. 4 is a picture prepared to understand the above explanation.

Fig. 4: Products with light reflection (left) and products without reflection (right)

In the case of the product on the left, a reflection of light occurs, so the left and bottom sides of the sticker, which must be identified, are not visible. In addition, reflection occurs in the center of the sticker, making it difficult to accurately detect the defective class. On the other hand, in the case of the product on the right, there is almost no light reflection, and especially the area of the sticker that detects the defective class is accurately illuminated.

Fig. 5: Normal determined by the inspection system

The cosmetic container, which is the subject of this study, undergoes the sticker attachment process after defective product. the injection process. If the position of the attached sticker is not correct or the sticker surface is damaged and an image different from the normal product is taken, the product must be passed through the discharge cylinder to a discharge part different from the normal product. Fig. 6 is a picture of the set-up system discriminating between a defective product and a normal product based on the above description. The normal class was not learned separately but was learned in three classes: sticker peeled off ('Peeled off'), wrinkled sticker ('Wrinkle'), and out of position ('Escape'). The reason for not learning the normal class is that the product's bad class features often include the characteristics of the normal class, so good performance was not obtained as a result of learning the normal class. Therefore, only images of defective products were defined and trained without learning the normal class. Of course, even if normal class and defective class were detected together, it could be classified through programming, but as a result of the experiment, the accuracy of the system was measured to be higher when only the defective class was learned.

Based on the defined class, Deep Learning Train was performed on the YOLOv4 model. 1 week from initial data learning to final inspection took time, and A total of 5 inspection courses were conducted and Before the experiment, it was trained using about 3,000. The reason for shortening the learning period was to examine whether accurate performance could be achieved even in a short learning period. As the production process focused on small-lot production of various products is highly demanded, it may be difficult to apply the requirements of various fields if a lot of learning time is required. Accordingly, to achieve maximum performance with minimal learning, the learning image and learning time were set to be small.

Fig. 6: Discharge of the identified defective product

4.2 Results

In the case of early learning, due to the difference between abnormal sample data and actual test data, a normal product was judged as defective (Over Detection) or a defective product was judged as normal into the cylinder. (Miss Detection). However, as the amount of learning of bad sample data and normal sample data gradually increased, it finally showed Precision 1, Recall 0.996. Table 3 is a tabular summary of the experimental results.

Table 3. Experimental Results.

	Total Test Count	Real Bad Product	Detected Bad Product	(FN)Over-Detection	(FP) Miss-Detection	Precision	Recall
1st Inspection	130	85	52	0	33	1.00	0.611
2nd Inspection	832	640	601	10	49	0.98	0.92
3rd Inspection	386	232	348	132	16	0.62	0.93
4th Inspection	152	106	105	0	1	1.00	0.99
Final Inspection	440	263	262	0	1	1.00	0.996

Since the bad class was learned in this experiment, FP corresponds to detecting the bad class in a good product. Conversely, FN is a case where a defective product cannot be detected as a defective class and is judged to be a normal product. You should look at the table with this in mind. For example, in the first inspection, 130 products were prepared, and 85 defective products were included. Here we detected 52 defective products. After checking the products that were identified, 33 defective products were identified among the products that were determined to be good. So there are 33 FNs. Conversely, among the products identified as defective products, there were 0 normal products. So FP will be zero. Now, according to the calculation formula, Precision = TP / (TP + FP), so Precision = 52 / (52 + 0) = 1 is derived. For Recall, Recall = TP / (TP + FN), which leads to Recall = 52 / (52 + 33) = 0.61. Viewing the table in this way will help you understand.

5 Conclusion

By using the designed real-time anomaly detection system, it was possible to reduce the number of people put into the inspection department at the site, and it was possible to obtain the result of accepting the fatigue of the existing workers. In addition, it has great significance in that it built a model that can be applied immediately in the field at a low cost.

Since this system is a model created based on Supervised Learning, collection of bad sample data is essential, and as shown in Table 1 of 4.2, if various data cannot be secured, good performance cannot be achieved. Accordingly, the goal is to review the GAN, [23] model and study an anomaly detection model, [24] that can be satisfied in the field even with Unsupervised Learning so that the data is produced by artificial intelligence and used as learning data.

This paper does not give specific suggestions on how to utilize the collected manufacturing data. There are infinite ways to utilize manufacturing data, and as a representative example, research to implement a predictive maintenance system using manufacturing data in the smart factory industry can be confirmed, [25]. As the number of products produced increases, the amount of data collected is vast, so big data-based platform research is also being prepared, [8], and it is attracting attention as data that can be used in various other systems. If raw data collected in small factories can be transmitted to Cloud, MES, or ERP servers using a simple system, it is judged that it will be helpful for applied research using this.

References:

[1] Son Young-Jin and Choi Hwan Young . “A Study on Smart Factory Introduction Cases and Sustainable Effect”, Journal of Practical Engineering Education (JPEE) , 2022, vol.14, no.1, pp. 127-136

[2] HOZDIĆ Elvis, “Smart factory for industry 4.0: A review”, International Journal of Modern Manufacturing Technologies, 2015, 7.1: 28-35.

[3] PAGNON William, “The 4th industrial revolution–A smart factory implementation guide”, International Journal of Advanced Robotics and Automation (IJARA), 2017, 2.2: 1-5.

[4] BOCHKOVSKIY Alexey and WANG Chien-Yao and LIAO Hong-Yuan Mark, “Yolov4: Optimal speed and accuracy of object detection”, arXiv preprint arXiv:2004.10934, 2020.

[5] CHALAPATHY Raghavendra and CHAWLA Sanjay, “Deep learning for anomaly detection: A survey”, arXiv preprint arXiv:1901.03407, 2019.

[6] KIM Chang-Min and CHUNG Kyungyong and PARK Roy C, “Anomaly detection model of mammography using YOLOv4-based histogram”, Personal and Ubiquitous Computing, 2021, 1-12.

[7] Tushar Jain and Meenu A, “Integration of Industries Through Computer Vision systems.”, International Journal of Information and Computation Technology, 2013 , ISSN 0974-2239 Volume 3, Number 9 , pp. 963-970

[8] SHIN Seung-Jun and WOO Jungyub and SEO Wonchul, “Developing a big data analytics platform architecture for smart factory”, Journal of Korea Multimedia Society, 2016, 19.8: 1516-1529.

[9] CUNNINGHAM and Pádraig CORD and Matthieu DELANY and Sarah Jane m , "Supervised learning". In: Machine learning techniques for multimedia , Springer, Berlin, Heidelberg, 2008. p. 21-49.

[10] Guansong Pang and Chunhua Shen and Longbing Cao and Anton Van Den Hengel , "Deep learning for anomaly detection: A review", ACM Computing Surveys (CSUR), 2021, 54.2: 1-38.

[11] OMAR Salima and NGADI Asri and JEBUR Hamid H, "Machine learning techniques for anomaly detection: an overview", International Journal of Computer Applications, 2013, 79.2.

[12] Ross Girshick and Jeff Donahue and Trevor Darrell and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, p. 580-587.

[13] Weisong Shi and Jie Cao and Quan Zhang and Youhuizi Li and Lanyu Xu, "Edge computing: Vision and challenges", IEEE internet of things journal, 2016, 3.5: 637-646.

[14] Blesson Varghese and Nan Wang and Sakil Barbhuiya and Peter Kilpatrick and Dimitrios S. Nikolopoulos, "Challenges and opportunities in edge computing", In: 2016 IEEE International Conference on Smart Cloud (SmartCloud), IEEE, 2016. p. 20-26.

[15] DAVIS Jesse and GOADRICH Markm "The relationship between Precision-Recall and ROC curves", In: Proceedings of the 23rd international conference on Machine learning, 2006, p. 233-240.

[16] Justin M. Johnson and Taghi M. Khoshgoftaar, "Survey on deep learning with class imbalance", Journal of Big Data, 2019, 6.1: 1-54.

[17] David M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", arXiv preprint arXiv:2010.16061, 2020.

[18] Oh Sewon and Jeong Jongpil and Park Jungsoo, "Design and implementation of smart factory system based on manufacturing data for cosmetic industry", The Journal of the Institute of Internet, Broadcasting and Communication, 2021, 21.1: 149-162.

[19] Kim, Yu-Sin and An-Seop Choi, "Analysis of the change of intensity distribution and luminous environment by reflector shape in indirect reflected LED luminaires", Journal of the Korean Institute of Illuminating and Electrical Installation Engineers 25.1 ,2011, p.: 9-17.

[20] Ryan Sun and Matthew B. Bouchard and Elizabeth M. C. Hillman, "SPLASSH: Open source software for camera-based high-speed, multispectral in-vivo optical image acquisition", Biomedical optics express, 2010, 1.2: 385-397.

[21] RHODY Harvey, "Lecture 10: Hough circle transform", Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 2005.

[22] BSubutai Ahmad and Alexander Lavin and Scott Purdy and Zuha Agha, "Unsupervised real-time anomaly detection for streaming data", Neurocomputing, 2017, 262: 134-147.

[23] Antonia Creswell and Tom White and Vincent Dumoulin and Kai Arulkumaran and Biswa Sengupta and Anil A. Bharath, "Generative adversarial networks: An overview", IEEE signal processing magazine, 2018, 35.1: 53-65.

[24] Subutai Ahmad and Alexander Lavin and Scott Purdy and Zuha Agha, "Unsupervised real-time anomaly detection for streaming data", Neurocomputing, 2017, 262: 134-147.

[25] CHUN Seung-Man and SUK Soo-Young, "Development and implementation of smart manufacturing big-data platform using opensource for failure prognostics and diagnosis technology of industrial robot", IEMEK Journal of Embedded Systems and Applications, 2019, 14.4: 187-195.

Conflicts of Interest

The author(s) declare no potential conflicts of interest concerning the research, authorship, or publication of this article.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

The author(s) contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of funding for research presented in a scientific article or scientific article itself

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2018-0-01417) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1060054).

Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)