

ARTICLE

YOLO-Based Damage Detection with StyleGAN3 Data Augmentation for Parcel Information-Recognition System

Seolhee Kim¹ and Sang-Duck Lee^{2,*}

¹Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, 16419, Republic of Korea

²Logistics System Research Division, Korea Railroad Research Institute, Uiwang, 16105, Republic of Korea

*Corresponding Author: Sang-Duck Lee. Email: sdlee@krii.re.kr

Received: 21 March 2024 Accepted: 11 June 2024 Published: 18 July 2024

ABSTRACT

Damage to parcels reduces customer satisfaction with delivery services and increases return-logistics costs. This can be prevented by detecting and addressing the damage before the parcels reach the customer. Consequently, various studies have been conducted on deep learning techniques related to the detection of parcel damage. This study proposes a deep learning-based damage detection method for various types of parcels. The method is intended to be part of a parcel information-recognition system that identifies the volume and shipping information of parcels, and determines whether they are damaged; this method is intended for use in the actual parcel-transportation process. For this purpose, 1) the study acquired image data in an environment simulating the actual parcel-transportation process, and 2) the training dataset was expanded based on StyleGAN3 with adaptive discriminator augmentation. Additionally, 3) a preliminary distinction was made between the appearance of parcels and their damage status to enhance the performance of the parcel damage detection model and analyze the causes of parcel damage. Finally, using the dataset constructed based on the proposed method, a damage type detection model was trained, and its mean average precision was confirmed. This model can improve customer satisfaction and reduce return costs for parcel delivery companies.

KEYWORDS

Parcel delivery service; damage detection; damage classification; data augmentation; generative adversarial network

1 Introduction

Consumers have recently shifted their consumption patterns to e-commerce and online channels, thereby accelerating the growth of parcel volumes [1,2]. Therefore, the number of international parcel volumes increased by 150%, from 6.4 billion in 2016 to 16.1 billion in 2022 [3], and this trend is expected to continue in the future. However, the expansion of parcel volumes also led to an increase in the frequency of parcel damage. According to a survey on the experiences of parcel damage, over 50% of respondents reported receiving damaged parcels for four consecutive years between 2019 and 2022 [4]. Continuous damage to parcels not only reduces consumer satisfaction with delivery services [5], but also leads to increased return-logistics costs, causing losses for delivery service providers. Numerous studies on deep learning-based parcel damage detection methods have been conducted to detect and



improve this issue during the parcel transportation process automatically. However, existing studies on parcel damage detection are limited in that they are only able to track the occurrence of damage without distinguishing the type of damage. This limitation prevents the precise identification of the cause of the problem. Therefore, this study proposes a method that distinguishes the types of parcel damage based on images of parcels, enabling the identification of both the causes and timing of damage occurrences.

The rationale for proposing a method based on image data to identify the causes of parcel damage is that damage from external factors primarily impacts the surfaces of parcels. By utilizing image-based classification of parcel damage types, as presented in this study, it is possible to detect damages early and address them effectively, thereby resolving the fundamental causes. Among these, 'contamination' and 'impact,' both major causes of damage, often result in widespread issues. 'contamination' can spread to other parcels, whereas 'impact' affects a wide area, typically impacting multiple parcels; thus, the early identification of the initial causes of parcel damage is crucial. In addition, damage to the areas of the parcel surface that display essential information for delivery, such as the invoice, can reduce the efficiency and reliability of the delivery service. Therefore, the parcel damage detection method proposed in this study, based on the image information of parcels, can prevent potential parcel damage issues that may occur during the parcel transportation process at a relatively low cost, thereby enhancing the reliability of delivery services.

The image information of a parcel can be acquired and analyzed at a relatively low cost, leading to various deep learning-based studies on damage detection. Previous studies have focused on classifying the damage status of parcel boxes, which constitute the largest proportion of parcel appearances [6,7]. However, to increase the parcel load in transit, recent parcel logistics processes use a variety of non-standard packaging materials such as Kraft, high-density polyethylene (HDPE), and low-density polyethylene (LDPE), which are smaller in volume. Thus, addressing damage in previous studies is difficult.

To address this issue, Kim et al. [6] utilized web crawling to collect images of various parcel appearances to train the damage detection model. However, the data collected through web crawling cannot reflect the real-time movement of parcels during transportation. In addition, this includes images from angles that are difficult to capture during the actual parcel transportation process, leading to a mismatch in the data collection environments. Furthermore, the inclusion of various objects in addition to parcels in a single image poses a limitation, potentially degrading the performance of the detection model.

To overcome these limitations, Kim [7] proposed a method to expand the amount of data based on a small number of high-quality datasets using data augmentation techniques. In particular, image processing techniques such as rotation and brightness changes were used to augment the data and improve the overfitting problem in the parcel damage status classification model. However, this approach cannot acquire new appearances of parcel images because it augments the amount of data by manipulating previously obtained images. In addition, random augmentation can lead to overfitting problems for specific data classes. Moreover, research utilizing this approach is limited in identifying the type of damage, making it difficult to determine the causes of damage.

To solve these problems, in this study, we simulate an actual parcel transportation environment and acquire images of various parcel appearances. Additionally, we expanded the dataset using StyleGAN3 with adaptive discriminator augmentation (ADA) to address the class imbalance in the detection model training dataset. Ultimately, the approach involves first detecting the appearance of parcels and classifying their damage status, followed by identifying the damage type, using a YOLOv5 model

trained on the aforementioned dataset for the real-time tracking of moving parcels. The proposed method involves first training the detection model using a dataset acquired in an environment that simulates the actual parcel transportation process, ensuring a high detection performance when applied to a real logistics terminal environment. Furthermore, grayscale images are acquired during data acquisition to reduce misrecognition that may occur with RGB images, contributing to the enhancement of the image generation focused on damage types and the performance of damage type detection. Second, expanding the dataset based on StyleGAN3 with ADA provides diversity in the training data and addresses the issue of class imbalance. Finally, by first detecting parcel appearances and classifying of their damage status, followed by the use of the YOLOv5-based damage detection process, we can enhance the performance of parcel damage detection and utilize it to analyze and improve the causes of parcel damage.

The remainder of this paper is organized as follows. [Chapter 2](#) explains generative adversarial networks (GANs) and object detection tasks. [Chapter 3](#) explains the dataset-construction method proposed in this study and the detection of the parcel damage type, followed by the detection of parcel types and classification of their damage status. [Chapter 4](#) verifies the performances of the constructed dataset and parcel damage detection model. Finally, [Chapter 5](#) discusses the practical and theoretical implications of the study. Finally, [Chapter 6](#) presents the conclusions of the study, its limitations, and the scope for further research.

2 Related Work

2.1 Generative Adversarial Networks

Goodfellow et al. [8] utilized GANs that learn through a structure in which generator and discriminator neural networks compete. In this process, the generator aims to create images similar to the original images, whereas the discriminator strives to distinguish between the generated and original images. However, the initially proposed GANs faced unstable training because of gradient vanishing, which makes the generation of high-quality images difficult. To address this issue, various studies have been conducted to modify GAN structures.

Radford et al. [9] utilized DCGAN that modifies the fully connected layers to convolutional neural networks (CNNs) to reflect the spatial characteristics of images and enables more stable training than GANs through architectural modifications, such as batch normalization and activation functions. In addition, it enables the manipulation of information within images through arithmetic operations on the latent vectors. However, this model has a limitation in that as the network deepens, the required computational load and the number of parameters increase significantly, making the generation of high-resolution images challenging.

Karras et al. [10] utilized StyleGAN2-ADA that improves model stability and the quality of generated images by maintaining a balance between the generator's image creation capabilities and the discriminator's ability to discern the generated images. Additionally, this model can be trained using a small amount of data without altering the loss function or network structure. ADA trains the discriminator using only augmented images, which include modifications such as rotation, scaling, color transformation, and RGB noise, and the generator also includes these augmented images in its training.

In previous StyleGAN models, aliasing occurred where the features of the image were fixed at specific coordinates. Karras et al. [11] utilized StyleGAN3 that addresses this problem by replacing traditional upsampling filters with filters based on the Kaiser window and designing a sequence

to prevent the addition of arbitrary high-frequency information when passing through nonlinear functions.

To generate a diverse range of damaged-parcel images, the damage characteristics must not be fixed to specific coordinates. Additionally, because the number of damaged parcels is low compared to that of normal parcels during the actual logistics process, generating images from a small number of damaged-parcel images is essential. Therefore, in this study, we utilize a StyleGAN3 model that employs ADA, which combines the generator structure of StyleGAN3 and the discriminator structure of StyleGAN2-ADA for data augmentation.

2.2 Object Detection Task

Object detection involves the extraction of localization information and the classification of objects. One-stage models carry out localization and classification simultaneously. By contrast, two-stage models proceed with localization and classification in two separate steps. One-stage models excel in terms of speed, whereas two-stage models offer advantages in terms of accuracy. Therefore, one-stage models are more suitable for detecting damage to parcels considering real-time movement.

One-stage models notably include the single-shot multi-box detector (SSD), RetinaNet, and You Only Look Once (YOLO). Liu et al. [12] demonstrated that an SSD predicts and detects candidate object areas within an input image by utilizing feature maps of various sizes. In this process, the size of the feature maps progressively decreased, enabling the detection of objects of various sizes. However, this approach can lead to performance degradation owing to difficulties in detecting small objects and class imbalances between the objects and background.

Lin et al. [13] demonstrated that RetinaNet was designed to address the issue of one-stage models having lower accuracy compared to two-stage models. It identifies the problem as class imbalance and uses focal loss, which assigns smaller weights to data that the model can easily distinguish, to resolve this issue. However, because it detects objects at multiple scales, its detection speed is limited.

Redmon et al. [14] demonstrated that YOLO is a model designed for fast object detection that is suitable for real-time object detection tasks. It divides an image into a grid and predicts the bounding boxes and class probabilities for each grid cell. YOLOv5 offers various model sizes and types, allowing the selection of a model that best fits the training objectives. Additionally, YOLOv5 introduces a classification model, YOLOv5-cls, which demonstrates faster speeds than other classification models [15]. Therefore, in this study, YOLOv5m is used to detect the appearance and type of damage in moving parcels, and YOLOv5s-cls is used to classify the damage status of the parcels. A detailed description of these models is provided in [Chapter 3](#).

3 Materials and Methods

3.1 Overview

3.1.1 Parcel Information-Recognition System

The parcel information-recognition system, which is the overarching structure of the parcel damage detection method proposed in this study, comprises three methods, as shown in [Fig. 1](#). For this purpose, a simulation of part of the actual parcel-transportation process was conducted, as shown in [Fig. 2](#). A photoelectric sensor was used to detect parcels moving continuously at 1 m/s, and this signal was relayed to a LiDAR sensor and an RGB camera to recognize information about the parcels. First, the three-dimensional volume-recognition method employs a LiDAR sensor (L515, Intel RealSense, USA) to measure the volume of the parcels. This volume information was used for stable loading

based on the size of the parcels during the last-mile stage. The barcode-recognition method employs the 4K RGB camera (BFS-U3-244S8M-C, Teledyne FLIR, USA) to recognize the barcode on the invoice attached to the parcel. This barcode number, which represents the invoice number, contains information such as the recipient, delivery location, and parcel type. This was used to arrange the loading of parcels according to the unloading sequence in the last-mile stage. Finally, the proposed damage detection method employs the 4K RGB camera (BFS-U3-244S8M-C, Teledyne FLIR, USA) to detect the damage type based on the appearance of the parcels. A detailed explanation of the parcel damage detection method follows.

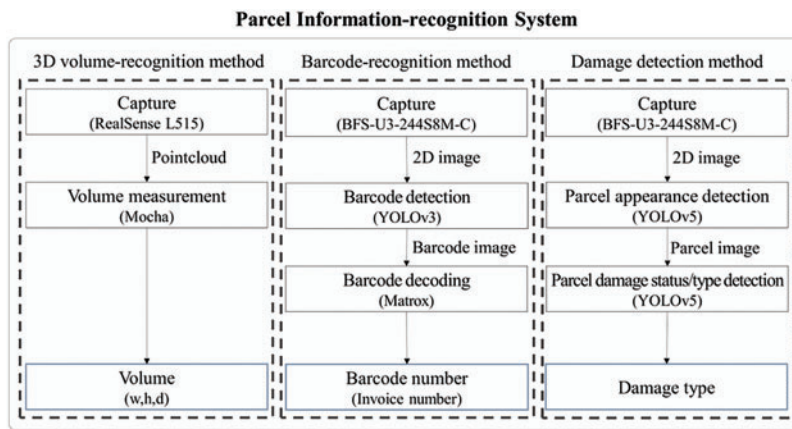


Figure 1: Flowchart of parcel information recognition system

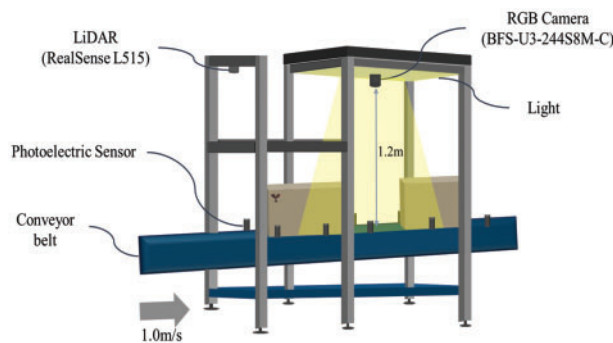


Figure 2: Example of parcel information-recognition system environment

3.1.2 Parcel Damage Detection Method

The entire process of the proposed parcel damage detection method comprises two stages, as illustrated in Fig. 3. Dataset construction was divided into the data acquisition, pre-processing, and image generation stages. First, in the data acquisition and pre-processing stage, original images of the moving normal and damaged parcels were acquired. Subsequently, using a pre-trained YOLOv5 detection model, the appearances of parcels are detected and classified into nine classes based on the acquired images. Next, the process involves cropping the parcel areas using the bounding box results from the parcel appearance detection and labeling them according to the damage status. Details of the data acquisition and pre-processing stages are described in Section 3.2.

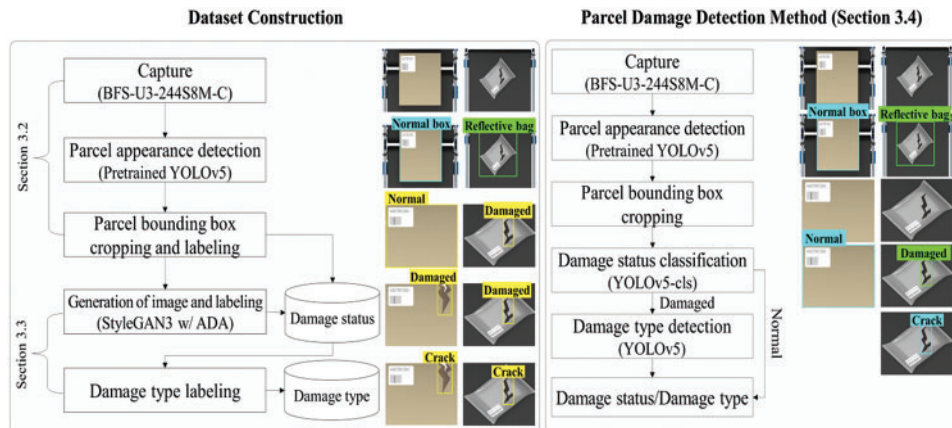


Figure 3: Flowchart of data construction and parcel damage detection method

In the image generation stage, new damaged images were created using the ADA with StyleGAN3, based on the original images of damaged parcels acquired in the previous stage, followed by labeling according to the damage status. Next, a dataset for the parcel damage status was constructed, including original images of normal and damaged parcels, as well as generated images of damaged parcels. Finally, labeling by the type of damage was performed for both the original and generated images of damaged parcels, ultimately constructing a dataset focused on the types of damage to parcels. Details of the image generation stage are described in [Section 3.3](#).

The method for detecting the type of damage involves receiving original images of the parcels using a camera. Next, it precedes the detection of the appearance of parcels and classifies the damage status, followed by the sequential detection of damage types. This approach addresses various types of damage beyond those proposed in this study and can enhance the performance of the damage type detection model. The details of the damage type detection method are described in [Section 3.4](#).

3.2 Data Acquisition and Pre-Processing

During the data acquisition and pre-processing stages of dataset construction, original images of parcels in transit and damaged parcels were acquired. Image collection based on web crawling in previous studies does not reflect the real-time movement of parcels and may include shooting points not implemented in the logistics environment, resulting in differences from the images acquired during the actual parcel transportation process. Owing to these differences, there is a limit to the performance improvements when applying detection models. Therefore, in this study, all images were acquired in the environment depicted in [Fig. 2](#), based on the scenario of a parcel logistics terminal in which parcels continuously move. The parcels moved on a conveyor belt at a speed of 1.0 m/s. To capture this, a 4K RGB camera with 4000 lm of lighting was installed on top of the conveyor belt at a height of 1.2 m.

Images of the parcels were acquired in grayscale. This minimizes the motion blur phenomenon, in which the boundary between the background and the object in the image becomes blurred during the shooting of fast-moving objects, which can degrade the performance of the detection models. In addition, grayscale images can shorten the computation time of the damage detection model for moving parcels and minimize the impact of the color of parcels on the classification and damage detection models. Before the image pre-processing stage, the acquired images were used to detect parcel appearances using the pre-trained YOLOv5 to extract the bounding boxes of the parcels and

parcel appearance classes. The pretraining of YOLOv5 was conducted based on the images acquired according to the parcel appearances listed in [Table 1](#).

Table 1: Classification of parcel image data

Structure	Material	Appearance (class)
Structured	Cardboard paper	Normal box
		Patterned box
		Paper tube
	Styrofoam	Styrofoam box
Unstructured	Kraft paper	Wrapped box
	HDPE	Black bag
	Paper	White bag
	LDPE	
	PET	Reflective bag
	N/A	Other

As shown in [Table 1](#), parcel types were classified into nine classes, categorizing the acquired images of parcels based on the structure, material, and appearance of the parcel. In this case, the parcel materials referenced packaging methods for various types of shipments, as presented in the standard terms and conditions for parcel delivery services by the Korea Fair Trade Commission [16]. In this context, structured parcels have standardized sizes and shapes, whereas unstructured parcels lack standard dimensions.

Based on the material and appearance of the parcels, they were categorized into nine types: Normal box, Patterned box, Paper tube, Styrofoam box, Wrapped box, Black bag, White bag, Reflective bag, and Others. A Normal box is a plain paper box without patterns, whereas a Patterned box is a paper box with designs. A Paper tube refers to a cylindrical container made of paper, and a Wrapped box refers to a box wrapped in smooth wrapping paper. Furthermore, the black bag is specifically made from black HDPE material, whereas the white bag is constructed from white paper and LDPE. The reflective bag is crafted from Polyethylene Terephthalate (PET) material that reflects like a mirror.

Because the types of packaging materials can vary depending on the packaging of unstructured parcels, materials that were not among the four most frequently used were classified into an “Others” class. Additionally, delivery service company trademarks, text, and barcodes, which all parcels may include, were not incorporated into the classification of parcel types. Examples of images from the nine classes used in the final classification in [Table 1](#) are shown in [Fig. 4](#).

During the image pre-processing stage, the bounding box areas of the parcels detected by the pre-trained YOLOv5m model were cropped, as shown in [Fig. 5](#). This process enabled the classification of parcels by appearance and the enlargement of parcel features that occupy relatively small areas within the images [17]. Subsequently, labeling was conducted according to the damage status based on images processed through the data acquisition and pre-processing stages, as described above.

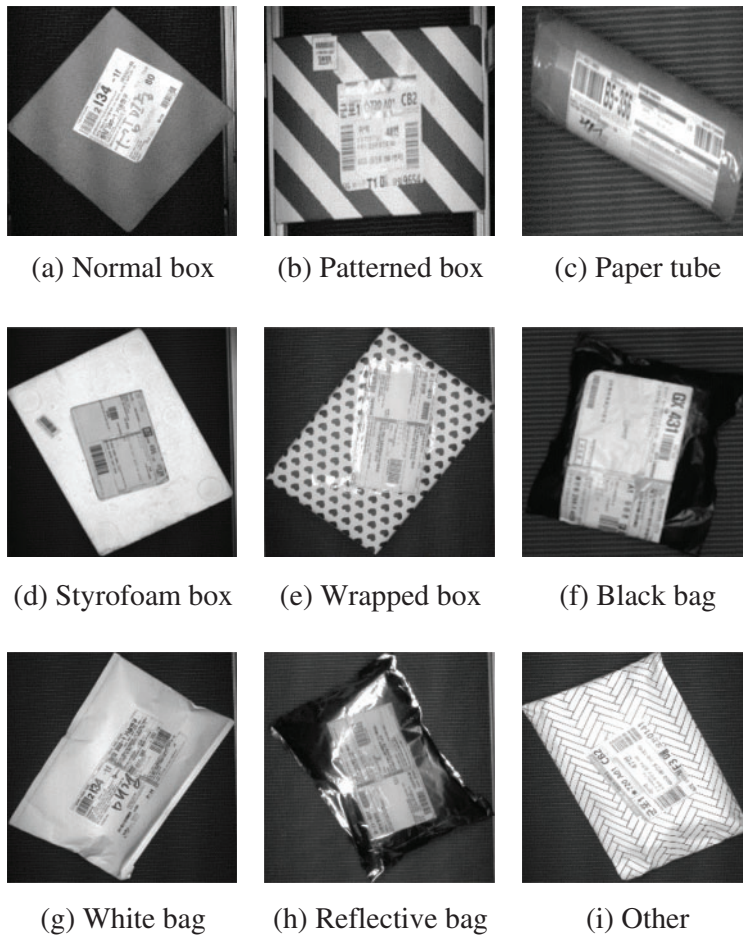


Figure 4: Example of parcel damage classification

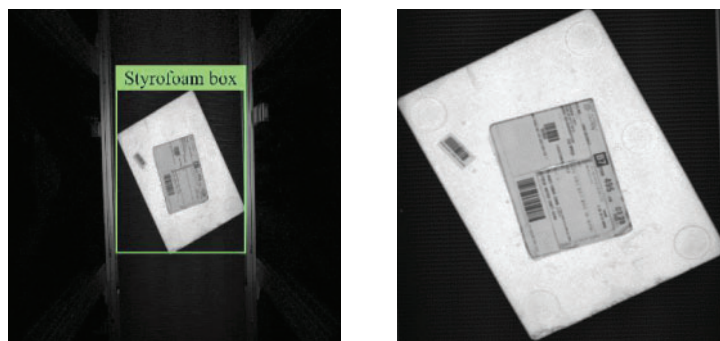

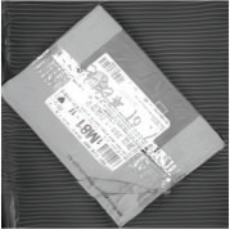
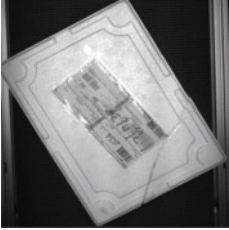





Figure 5: Example of comparison of images before and after detection and cropping

3.3 Image Generation

In the image generation stage, images processed through the data acquisition and pre-processing steps described in Section 3.2 are classified based on their damage status. However, the number of images of damaged parcels that can be acquired during the actual parcel transportation process is limited compared to the number of images of normal parcels. This results in a class imbalance between damaged and normal parcels. To address this imbalance, images of damaged parcels were generated using StyleGAN3 with ADA, utilizing the images of damaged parcels acquired in Section 3.2. Subsequently, a dataset of the parcel damage status was constructed, including both the generated images of damaged parcels and original images classified according to the parcel damage status, as shown in Table 2. The images generated in Table 2 result from training StyleGAN3 with ADA on 34, 24, and 22 original images for a single appearance of parcels per class.

Table 2: Example images of damaged parcels: (a) Normal box, (b) Styrofoam box, (c) White bag

Class	Original image	StyleGAN3 w/ADA
(a)		
(b)		
(c)		

Next, to construct a dataset on the parcel damage types, images of damaged parcels were labeled within the dataset on the parcel damage status. The classes of parcel damage types were categorized based on the types of damage that can occur in structured and unstructured parcels. Table 3 describes the types of parcel damage classes. “crack” refers to cases where cracks appear on the parcel’s surface, typically caused by excessive pressure, scratching from sharp objects, or the material’s expansion and contraction owing to temperature changes. “dent” indicates the parcel surface being twisted owing to external impacts, commonly occurring from collisions with heavy objects, falls, or impacts with other parcels. “hole” denotes situations where a hole is punctured in the parcel surface, revealing the interior, usually owing to collisions with sharp objects or strong pressure. “split” describes the parcel

being torn, allowing some of the contents to spill out, which happens when parcels are excessively stacked, mishandled, or when the packaging material weakens owing to contamination.

Table 3: Description of parcel damage types

Damage types	Description
Crack	Cracks on the surface of the parcel box
Dent	Twist on the surface of the parcel box
Hole	Holes on the parcel box surface reveal the interior
Split	Parcel-box surface torn with some contents exposed

Structured parcels, which typically have thicker packaging materials, are less likely to experience “split” damage, even when coming into contact with sharp objects. Unstructured parcels composed of more flexible materials are less likely to experience “dent” damage. Therefore, structured parcels were assigned “crack,” “dent,” and “hole” as damage types, whereas unstructured parcels were assigned “crack,” “hole,” and “split” as their damage types [18].

Through this process, a dataset of parcel damage images, including each damage type class, was constructed. The datasets constructed in Sections 3.2 and 3.3 were used as training data for models to classify the damage status of parcels and detect the types of parcel damage, respectively.

3.4 Parcel Damage Detection Method

In the process of detecting damage types, grayscale images of parcels were first analyzed to identify the parcel appearances and classify the damage status before detecting the specific types of damage. This preliminary step was crucial because the appearance of the parcel influenced its shape and durability, which in turn affected the type of damage that could occur. Therefore, the preceding process allowed for the accurate detection of damage types and enabled optimized damage management based on the damage type. In addition, this method allowed for the exception handling of damage types not included in the classes of crack, dent, hole, and split proposed in this study, thereby enabling responses to various types of damage. Furthermore, instead of uniformly detecting six classes (four types of damage, other, and normal) as parcel damage types, this approach ensured high performance by distinguishing between binary classification through damage status classification and the classification of the four types of damage.

First, parcel images were collected using a camera and input into a parcel appearance detection model using the pre-trained YOLOv5m. They were classified into nine parcel appearances, as detailed in Section 3.2, and the parcel areas were then cropped within the images. The cropped parcel images were then fed into a parcel damage status classification model for each parcel appearance using YOLOv5s-cls. If the determined that the parcel was normal, this status was outputted as the final result. Conversely, if the parcel was damaged, the damaged parcel image was input into a parcel damage type detection model for each type of parcel, using YOLOv5m to identify the damage type. Finally, the model detected and output the types of damage, including crack, dent, hole, and split, from the input damaged-parcel images.

The overall structure of the YOLOv5 detection model is depicted in Fig. 6 [19]. It comprises three parts: the backbone, neck, and head. The backbone serves as the main body of the network and utilizes the CSP-Darknet [20] architecture. The neck connects the backbone to the head and adopts the PANet

[21] structure. Finally, the head is responsible for generating the final prediction results and comprises three conv2d layers. The overall structure of the YOLOv5-cls classification model is shown in Fig. 7. It comprises two parts: the backbone and head, with the backbone part using the CSP-Darknet structure, similar to the detection model. The head is designed for classification and consists of pooling dropout and linear layers.

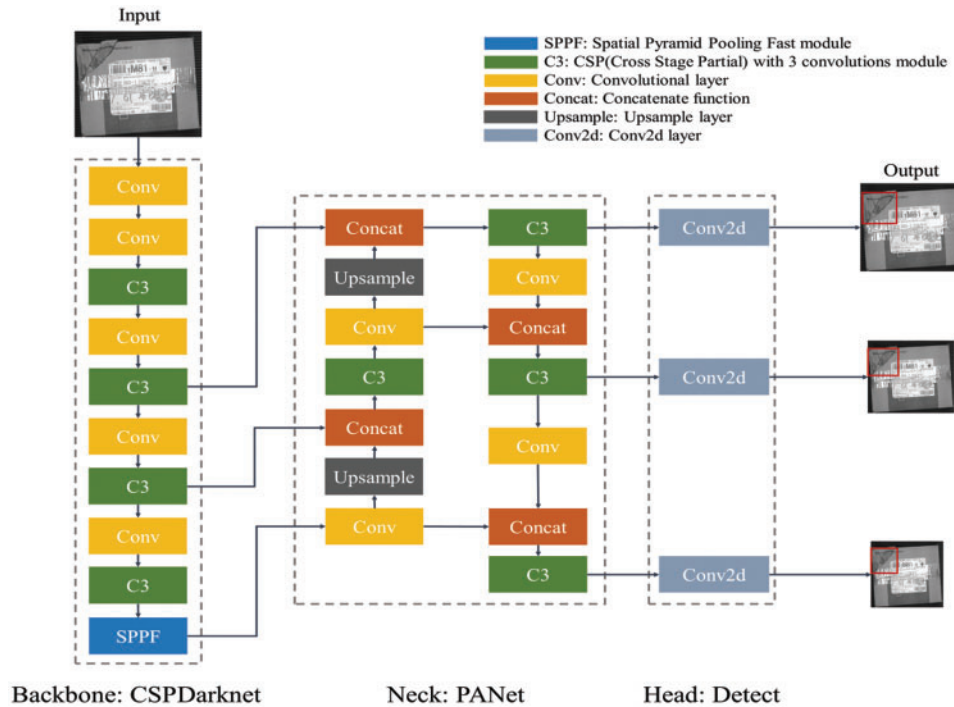


Figure 6: YOLOv5 architecture

4 Experiments

To validate the proposed parcel damage detection method, original images of both normal and damaged parcels were acquired under the conditions described in Section 3.2. Furthermore, through the image pre-processing and image generation processes described in Sections 3.2 and 3.3, respectively, two datasets were constructed for training the damage type detection model, which followed the detection of parcel appearances and classification of damage status. These experiments were based on a scenario that detected damage to parcels transported on a conveyor within the hub terminal of a parcel delivery company. Image generation and damage detection model training was conducted on a Windows 11 OS environment, utilizing an AMD Ryzen Threadripper PRO 5975WX 32-Core CPU and two NVIDIA RTX A6000 GPUs. Additionally, CUDA 12.0, cuDNN 8.6.0, Python 3.10, and the PyTorch framework were used. The image size used in this study was 256×256 , and the hyperparameters for each step are shown in Table 4.

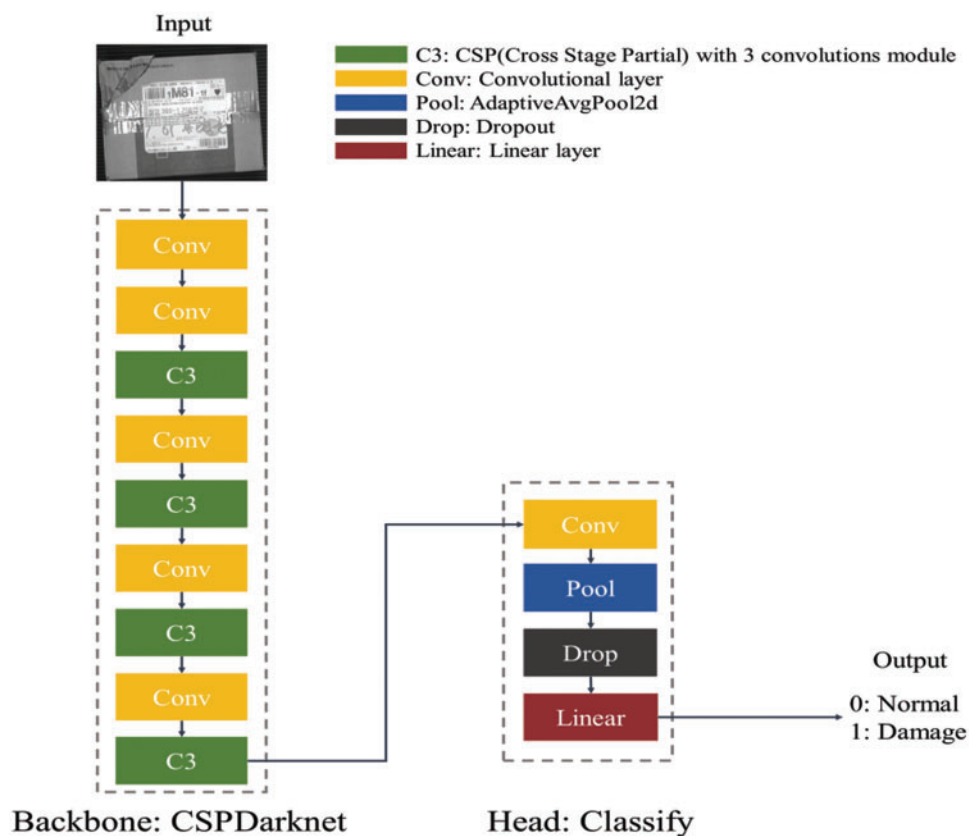


Figure 7: YOLOv5-cls architecture

Table 4: Image generation and damage type detection stage hyperparameters

Types	Image generation		Damage type detection		
			Parcel appearances detection	Parcel damage classification	Parcel damage type detection
Epoch	1500	200	100	200	
Batch size	16	32	64	32	
Learning rate	Gen.	2.5e-3	1e-2		
	Dis.	2e-3			
Optimizer	Adam		Adam		
Loss function	Softplus		Cross entropy		

4.1 Image-Acquisition and Pre-Processing

Parcel images were acquired based on the nine parcel appearance classes described in Section 3.2. The environment for acquiring the images was the same as that presented in Section 3.2. The

performance of the trained model varied depending on whether the acquired image data reflected the actual parcel transportation process in a logistics terminal. Therefore, in this study, conveyor belts were used to transport parcels in real logistics terminal environments, and images of the parcels moving on these conveyor belts were acquired. The dataset for pretraining, classified by the appearance of parcels, totaled 10,987 images. This dataset features a variety of parcel orientations, parcel positions on the conveyor belt, and invoices from different delivery service providers. The numbers of images for each parcel class are listed in [Table 5](#).

Table 5: Numbers of images for each parcel class

Class	Number of parcel images
Normal box	4332
Patterned box	500
Paper tube	708
Styrofoam box	668
Wrapped box	816
Black bag	391
White bag	1124
Reflective bag	1623
Other	825

[Fig. 8](#) shows examples of the images generated before and after pre-processing. [Fig. 8a](#) shows the original parcel image without pre-processing, and [Fig. 8b](#) shows an image generated by a model trained using the original image. Furthermore, [Fig. 8c](#) depicts the original parcel image after applying the pre-processing steps, and [Fig. 8d](#) illustrates an image generated by a model trained on the pre-processed image. The images generated in [Fig. 8](#) resulted from training StyleGAN3 with ADA using 26 original images of a single parcel appearance.

As shown in [Fig. 8a,b](#), images generated by models trained on parcel images without applying pre-processing methods retained the shape of the surrounding environment of the parcel, such as the conveyor belt, but failed to accurately reflect the appearance of the parcel and its damage. Conversely, as shown in the example in [Fig. 8c](#), images generated by models trained on parcel images that underwent pre-processing, such as those depicted in [Fig. 8d](#), accurately reflected both the appearance of the parcel and the damage type. Additionally, these models generated images that included small-sized information, such as the terminal number on the parcel's attached invoice. Consequently, by applying the proposed image pre-processing methods, the characteristics of parcel damage on a conveyor belt could be accurately captured and detected despite variations in the position of the parcel, and images that reflect these features could be generated.

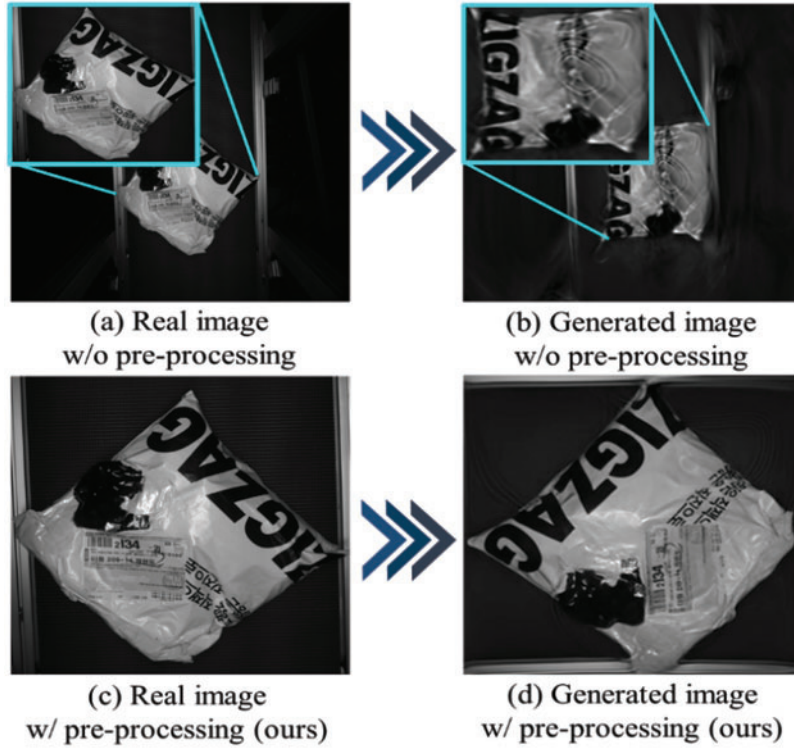


Figure 8: Examples of generated images before and after pre-processing

4.2 Image Generation

To verify the image generation performance of StyleGAN3 with ADA, as described in [Section 3.3](#), we performed both qualitative and quantitative evaluations of images generated using four adversarial generative models: DCGAN, StarGAN, StyleGAN3, and StyleGAN3 with ADA. DCGAN preserves spatial information by applying CNNs to the structure of vanilla GAN. StarGAN is designed for image-to-image translation and helps transform image attributes using a single model by encoding target domain labels. StyleGAN3 enables style modulation by incorporating intermediate layers into the network structure and contributes to high-resolution image generation and artifact reduction. The proposed StyleGAN3 with ADA increases the stability of the model and improves the quality of the generated images by balancing the image generation capability of the generator and the image recognition capability of the discriminator more effectively than StyleGAN3. Thus, considering the unique contributions of each model, we compared and analyzed the generated videos using the inception score (IS) and Fréchet inception distance (FID) evaluation metrics. To corroborate the effectiveness of the proposed method for detecting damage types, this study also examined the capabilities of the IS in evaluating the quality and diversity of generated images [22], and the FID score in assessing the similarity by calculating the Fréchet distance between the original and generated images [23]. These assessments are derived using [Eqs. \(1\)](#) and [\(2\)](#):

$$\text{IS}(G) = \exp(E_x D_{KL}(p(y|x) \| p(y))) \quad (1)$$

$$\text{FID} = \|\mu_1 - \mu_2\|_2^2 - \text{Tr}(\Sigma_1 + \Sigma_2 - 2\Sigma_1\Sigma_2) \quad (2)$$

In Eq. (1), G represents the generator, E is the expectation, and D_{KL} denotes the Kullback-Leibler divergence. The IS can be calculated by predicting the class probabilities for the images generated using a pre-trained Inception-v3 classification model. This process uses the expected value of the Kullback-Leibler divergence to measure the difference between the conditional and marginal probability distributions. A higher IS value indicates better quality and diversity of the generated images. In Eq. (2), μ_1 and μ_2 represent the means of the probability distributions of the generated and original images, and Σ_1 and Σ_2 represent the covariance matrices of their image probability distributions, respectively. The FID calculated the mean and covariance of the feature maps extracted from the intermediate layers of the images using the Inception-v3 classification model. The image generation results by each model are presented in Table 6, and the results of each model according to the evaluation metrics are listed in Table 7. The images generated in Table 6, representing structured and unstructured parcels, resulted from training StyleGAN3 with ADA using 34 and 25 original images of a single parcel appearance, respectively.

Table 6: Comparison of damaged-parcel images generated by models: (a) Structured, (b) Unstructured

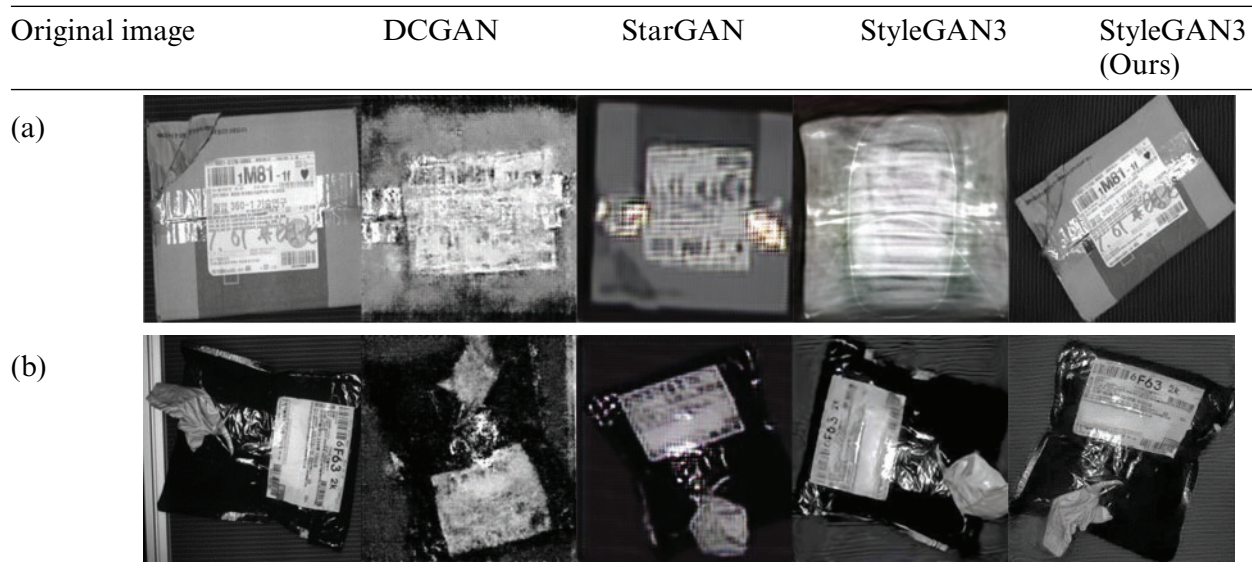


Table 7: Quantitative performance evaluation of generated parcel images

Methods	IS \uparrow ($\times 10^1$)	FID \downarrow
DCGAN	15.59	247.6
StarGAN	17.50	207.8
StyleGAN3	22.50	134.39
StyleGAN3 w/ADA	23.59	58.50

As indicated in Table 6, the images generated by the DCGAN model suffer from low quality owing to complexity limitations, thereby enabling the accurate reflection of the appearances of parcels and damage types. Images created with the StarGAN model, an image-to-image translation technique, reflect the parcel and damage types but include noise values not present in the original images. Images

generated by the StyleGAN3 model reflect the appearance of parcels and damage types but fail to capture detailed features such as invoices. In contrast, images generated by StyleGAN3 with ADA clearly reflect both the appearance of parcels and damage types, enabling the accurate generation of images. Therefore, to construct a training dataset for the parcel damage status classification and type detection, images generated using the StyleGAN3 with ADA were added to the dataset.

As shown in Table 7, the IS for StyleGAN3 with ADA was 23.59, which was higher than those for the DCGAN, StarGAN, and StyleGAN3 models. Furthermore, the FID score for StyleGAN3 with ADA was 58.50, which was lower than those for the DCGAN, StarGAN, and StyleGAN3 models. This demonstrated that when trained on the same damaged-parcel images, StyleGAN3 with ADA outperformed the other three models in terms of diversity and produced higher-quality images.

To establish datasets for the damage status of parcels and damage types, we set four classes of parcel damage, as explained in Section 3.3. The composition of each dataset was as follows. The damage status dataset comprised 6823 original images of normal parcels and 4164 original images of damaged parcels as well as 2776 generated images of damaged parcels, totaling 13,763 images. The damage type dataset included 4164 original images, with 1041 images for each damage type class from the damage status dataset, and an additional 2776 images, with 694 generated images per class using StyleGAN3 with ADA, totaling 6940 images. The detailed compositions are listed in Table 8. The datasets were split into training and validation data at an 8:2 ratio, with the validation dataset comprising only the original images and excluding the generated images.

Table 8: Construction of damage status and type dataset

		Normal	Damaged
Train	Original	5458	2776
	Generated	0	2776
Val.	Original	1365	1388
Total		6823	6940

4.3 Parcel Damage Type Detection

To validate the performance of the proposed damage type detection method, the performance of the damage type detection model was compared based on whether the detection of parcel appearances and classification of the parcel damage status was conducted beforehand. The mean average precision (mAP) was selected as the evaluation metric for the damage type detection model, and the results are presented in Table 9. Here, mAP is derived by averaging the area under the precision-recall curve for each class using Eqs. (3) and (4):

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (3)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (4)$$

TP represents true positives, FP represents false positives, TN represents true negatives, and FN represents false negatives. Precision is defined according to Eq. (3), which represents the ratio of samples correctly identified as positive to all samples predicted as positive by the model. Recall, defined according to Eq. (4), represents the ratio of samples correctly identified as positive by the model to

all actual positive classes. Because precision and recall are inversely related, a higher threshold for the intersection over union (IoU) applies a stricter criterion for identifying positives, which can increase precision, but may lower recall by misclassifying actual positives as negatives. Therefore, setting an appropriate threshold is necessary. In this study, the IoU threshold for YOLOv5m was set to 0.5, which was defined as true positive in the Pascal VOC object detection competition [24].

Table 9: Detection result of YOLOv5 using various anchor box settings

Settings	Anchor scales	Aspect ratios	mAP(%)
1 Scale	62×45	0.77	0.973
3 Scales	{ 10×13 , 62×45 , 373×326 }	0.77, 1.38, 1.14	0.975
9 Scales (default)	{ 10×13 , 16×30 , 33×23 }, { 30×61 , 62×45 , 59×119 }, { 116×90 , 156×198 , 373×326 }	{0.77, 0.53, 1.43}, {0.49, 1.38, 0.50}, {1.29, 0.79, 1.14}	0.98

Additionally, anchor boxes serve as initial reference points for predicting the location and size of objects in an object detection model. They are set with various scales and aspect ratios, providing a baseline for predicting the bounding boxes of objects. By using anchor boxes that match the actual scale and ratio of the objects, the model can predict bounding boxes more accurately, which contributes to the optimization of the model's mAP [25]. Generally, small scales of anchor boxes are used for detecting small objects, whereas larger scales are used for larger objects. Therefore, this study compared the performance according to different scales. The comparison group used the minimum, medium, and maximum values of the anchor boxes adopted by YOLOv5m.

When using only the median anchor box size of 62×45 in the YOLOv5m model, the mAP is 0.97. In contrast, when using the minimum, median, and maximum values of the anchor boxes, the mAP increased to 0.975, demonstrating improved detection performance compared to using just one anchor box size. This improvement indicates that the model can detect areas with significantly large or small damages more effectively. Additionally, when employing the 9 scale anchor boxes available in the YOLOv5m model, the mAP reached 0.99, allowing the model to effectively detect small, medium, and large objects. This capability significantly enhances object detection performance in complex visual environments. Therefore, in real-world logistics industry settings where objects of various sizes exist, it is crucial to use a combination of multiple anchor box scales to enhance the utility of the YOLOv5m model. This approach not only increases mAP but also maximizes detection accuracy and reliability.

As shown in Table 10, the mAP of the parcel damage type detection model without the prior detection of parcel appearances and classification of parcel damage status was 0.94, whereas the mAP of the model with these tasks conducted beforehand was 0.98, indicating an improvement in mAP performance across all classes. This suggests that the use of damage classification and type detection models tailored to the appearance each parcel can enhance the performance of parcel damage type detection. Moreover, when applied to actual logistics terminals, this implies that delivery service providers can enable optimized damage management according to the damage type. Additionally, the computational speed of the parcel damage type detection model, after conducting prior detection of parcel appearances and classification of parcel damage status, averaged 41.6 ms per parcel appearance. Thus, when applying the method proposed in this study to real logistics terminal environments, the number of parcel images that can be processed per second is approximately 24.

Table 10: Performance evaluation of damage type detection considering the precedence of parcel appearance detection and damage status classification

Methods	mAP per class (%)		Overall mAP (%)
Preliminary steps omitted –parcel appearance detection, –damage status classification	Crack	0.91	0.94
	Dent	0.97	
	Hole	0.92	
	Split	0.96	
Preliminary steps included +parcel appearance detection, +damage status classification	Crack	0.98	0.98
	Dent	0.99	
	Hole	0.97	
	Split	0.99	

5 Discussion

This study proposes a novel method for detecting various types of damage in parcels, utilizing an expanded dataset enhanced by StyleGAN3 with ADA. Validated by the YOLOv5m model, this method demonstrated high efficiency as evidenced by its mAP score. This chapter discusses the practical applicability of the proposed method and its implications for future research.

The developed method can be readily applied in the parcel delivery industry, enhancing the efficiency and reliability of damage detection processes. By integrating this system into existing parcel sorting workflows, it becomes possible to accurately assess the condition of parcels in real-time as they transit through hub terminals. This capability enables the early detection of damages, thereby reducing costs associated with parcel returns and minimizing customer dissatisfaction. Furthermore, in logistics warehouses, this system can be employed to inspect incoming and outgoing parcels, ensuring they meet quality standards and maintain the integrity of the supply chain.

Future research needs to expand the dataset to include scenarios under different environmental conditions, which will help improve the robustness and reliability of the deep learning models. In addition, if both the type and location of damage can be quantitatively measured in parcel damage detection, it is expected that clear criteria can be provided to parcel companies in insurance claims and compensation processes. Therefore, research will be conducted to quantify parcel damage based on damage segmentation.

6 Conclusion

In this study, a method capable of detecting various types of damage to different parcel appearances was proposed, and to implement this method, images were acquired in an environment that simulated the actual parcel transportation process. In addition, the dataset was expanded using StyleGAN3 with ADA. Finally, a method for detecting four types of damage (crack, dent, hole, and split) was implemented, which enabled the real-time detection of parcel appearances and classification of the damage status in moving parcels. To validate this, the mAP of the YOLOv5m model, which was trained using the dataset constructed in the aforementioned manner, was examined. The proposed method offers several key advantages over existing parcel damage detection techniques:

1) Improved data reliability: By addressing the issue of irrelevant objects in images collected through web crawling, which are unrelated to the logistics field, and resolving discrepancies between these images and actual logistics site conditions, our approach enhances the reliability of the data used for training and testing the model.

2) Enhanced dataset diversity: We expanded the dataset using StyleGAN3 with ADA, which enhances the diversity of the training data. Furthermore, augmenting data from minority classes within the dataset helps to address the issue of class imbalance, leading to more robust and accurate damage detection.

3) Detailed damage type classification: Unlike previous studies that only classify whether a parcel is damaged or not, our proposed method can identify parcel damages classified into four specific types (crack, dent, hole, and split). This detailed classification enables a more comprehensive analysis of the causes of parcel damage, contributing to increased customer satisfaction and reduced return costs for delivery service companies.

The proposed method has the potential to significantly improve the efficiency and accuracy of parcel damage detection in the logistics industry. By accurately identifying and classifying different types of parcel damage, delivery service companies can take proactive measures to address the root causes of damage, ultimately reducing costs associated with returns and enhancing customer satisfaction.

However, this study has some limitations that should be addressed in future research. Due to time and cost constraints, the amount of data utilized was limited. Collecting a larger volume of data on damaged parcels could significantly enhance the model's performance and generalizability. Additionally, analyzing the cost-effectiveness of the parcel information recognition system would provide valuable insights into the feasibility of implementing this system at the hub terminals of actual delivery companies. This analysis could serve as a benchmark for measuring the return on investment against the system's deployment.

Future research should focus on expanding the dataset to include a wider variety of parcel types, damage severities, and environmental conditions. Furthermore, exploring the integration of additional data sources, such as sensor data or customer feedback, could provide a more comprehensive understanding of parcel damage and its impact on the logistics process.

In conclusion, the proposed method offers a promising approach to improve parcel damage detection, with the potential to benefit both customers and delivery service companies. By addressing the limitations identified in this study and continuing to refine the model's performance, future research can contribute to the development of more efficient, accurate, and cost-effective parcel damage detection systems for the logistics industry.

Acknowledgement: This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) and the Korea Evaluation Institute of Industrial Technology (KEIT).

Funding Statement: This work was supported by a Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure, and Transport (Grant 1615013176) (<https://www.kaia.re.kr/eng/main.do>, accessed on 01/06/2024). This work was also supported by a Korea Evaluation Institute of Industrial Technology (KEIT) grant funded by the Korean Government (MOTIE) (141518499) (<https://www.keit.re.kr/index.es?sid=a2>, accessed on 01/06/2024).

Author Contributions: The authors confirm their contributions to the paper as follows: study conception and design: Seolhee Kim and Sang-Duck Lee; data collection: Seolhee Kim; analysis and interpretation of results: Seolhee Kim and Sang-Duck Lee; draft manuscript preparation: Seolhee Kim. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data not available due to ethical restrictions. The dataset cannot be shared owing to the inclusion of sensitive personal information.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] B. K. Kim, J. S. Park, H. B. Ryu, and D. Nam, "A study on the estimation and prediction of daily delivery service volume in the metropolitan area: Focuses on the case of Gyeonggi-Do," *Korean J. Logist.*, vol. 31, no. 2, pp. 47–58, 2023.
- [2] S. Han, J. U. Won, S. L. Seok, and E. Sampil, "The study on the efficiency of parcels unloading robot at delivery logistics terminal," *Korean J. Logist.*, vol. 28, no. 6, pp. 1–11, 2020.
- [3] "Pitney bowes parcel shipping index," Accessed: Jan. 12, 2023. [Online]. Available: <https://www.pitneybowes.com/us/shipping-index.html>
- [4] "Consumer home delivery review 2022/23," Accessed: Jan. 12, 2023. [Online]. Available: <https://www.imrg.org/insight/consumer-home-delivery-review-2022-23/>
- [5] J. H. Yang, "The relationship between relational benefits and loyalty of customers in courier services," *Korean J. Logist.*, vol. 19, no. 3, pp. 83–103, 2011. doi: [10.15735/klis](https://doi.org/10.15735/klis).
- [6] E. K. Kim, S. H. Kim, H. S. Sin, S. Y. Kim, and B. Lee, "Classification for the breakage of the package boxes using a deep learning network," in *Proc. KIBME*, Seoul, Republic of Korea, Jun. 20–22, 2022, pp. 250–253.
- [7] M. Kim, "Improvement of recognition performance through refinement of parcel damage classification algorithm based on CNN," Ph.D. dissertation, Univ. Sci. Technol., Daejeon, 2022.
- [8] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020. doi: [10.1145/3422622](https://doi.org/10.1145/3422622).
- [9] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.
- [10] T. Karras *et al.*, "Training generative adversarial networks with limited data," in *Proc. NeurIPS 2020*, Dec. 6–12, 2020, pp. 12104–12114.
- [11] T. Karras *et al.*, "Alias-free generative adversarial networks," in *Proc. NeurIPS 2021*, Dec. 6–14, 2021, pp. 852–863.
- [12] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *Computer Vision–ECCV 2016*, Amsterdam, The Netherlands, Oct. 11–14, 2016, pp. 21–37.
- [13] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) 2017*, Venice, Italy, Oct. 22–29, 2017, pp. 2980–2988.
- [14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit. Conf. (CVPR)*, Las Vegas, NV, USA, Jun. 26–Jul. 1, 2016, pp. 779–788. doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [15] "YOLOv5," 2020. Accessed: Aug. 30, 2023. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [16] "Packaging a Parcel," Accessed: Sep. 4, 2023. [Online]. Available: https://www.easylaw.go.kr/CSP/CnpClsMain.laf?popMenu=ov&csmSeq=663&ccfNo=2&cciNo=1&cnpClsNo=1&search_put=
- [17] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit. Conf. (CVPR)*, Boston, MA, USA, Jun. 7–12, 2015, pp. 815–823.

- [18] Z. Wang, J. Gao, Q. Zeng, and Y. Sun, "Multitype damage detection of container using CNN based on transfer learning," *Math. Probl. Eng.*, vol. 2021, no. 10, pp. 1–12, 2021. doi: [10.1155/2021/3839800](https://doi.org/10.1155/2021/3839800).
- [19] "Ultralytics YOLOv5 architecture," Accessed: Feb. 7, 2024. [Online]. Available: https://docs.ultralytics.com/ko/yolov5/tutorials/architecture_description/
- [20] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
- [21] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "PANet: Few-shot image semantic segmentation with prototype alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, Republic of Korea, Oct. 27–Nov. 2, 2019, pp. 9197–9206.
- [22] T. Salimans *et al.*, "Improved techniques for training GANs," in *Proc. NeurIPS 2016*, Barcelona, Spain, Dec. 5–10, 2016, pp. 2234–2242.
- [23] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. NeurIPS 2017*, Long Beach, CA, USA, Dec. 4–9, 2017, pp. 6629–6640.
- [24] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010. doi: [10.1007/s11263-009-0275-4](https://doi.org/10.1007/s11263-009-0275-4).
- [25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NeurIPS 2015*, Montreal, Canada, Dec. 7–12, 2015.