Enhancement of Image Classification Using Transfer Learning and GAN-Based Synthetic Data Augmentation

Subhajit Chatterjee [1], Debapriya Hazra [1], Yung-Cheol Byun [1,]* and Yong-Woon Kim [2]

[1] Department of Computer Engineering, Jeju National University, Jeju 63243, Korea; subhajitchatterjee@stu.jejunu.ac.kr (S.C.); debapriyah@jejunu.ac.kr (D.H.)
[2] Centre for Digital Innovation, CHRIST University (Deemed to be University), Bengaluru 560029, Karnataka, India; jonathan.kim@christuniversity.in
[*] Correspondence: ycb@jejunu.ac.kr

Abstract: Plastic bottle recycling has a crucial role in environmental degradation and protection. Position and background should be the same to classify plastic bottles on a conveyor belt. The manual detection of plastic bottles is time consuming and leads to human error. Hence, the automatic classification of plastic bottles using deep learning techniques can assist with the more accurate results and reduce cost. To achieve a considerably good result using the DL model, we need a large volume of data to train. We propose a GAN-based model to generate synthetic images similar to the original. To improve the image synthesis quality with less training time and decrease the chances of mode collapse, we propose a modified lightweight-GAN model, which consists of a generator and a discriminator with an auto-encoding feature to capture essential parts of the input image and to encourage the generator to produce a wide range of real data. Then a newly designed weighted average ensemble model based on two pre-trained models, inceptionV3 and xception, to classify transparent plastic bottles obtains an improved classification accuracy of 99.06%.

Keywords: deep learning; generative adversarial networks; image classification; transfer learning; plastic bottle

MSC: 68U10

1. Introduction

Due to flexibility in terms of cost, light weight, processing, and ease of carrying, plastic bottles are the most widely used material in daily life and industrial fields. Every day, tons of plastic bottles are dumped as waste, and in addition, toxic, hazardous materials in the trash are polluting the environment day by day [1]. An essential strategy for dealing with this issue is the recycling process. Recycling plastic bottles can be used further in new products, automobiles, textile, etc. Plastic bottle recycling has recently emerged as a significant part of the plastic bottle industry, potentially saving fossil fuels while simultaneously lowering greenhouse gas emissions [2].

The recycling task involves a lot of labor cost, and the DL approach helps in the way to automatically classify waste plastic bottles for recycling tasks [3]. Much research has been conducted to find a category of cost-effective PET bottle classifiers. PET bottles can be divided into several categories based on chemical resins, transparency, and color [4]. PET plastic bottles have the highest recycling values compared to other plastic bottles. The Ministry of Environment announced on 5 February that it would start a pilot project for the separation and disposal of transparent plastic bottles from this month. At the beginning of this month, five regions were phased out individually, including Busan, Cheonan (Chungnam), Gimhae (Gyeongnam), Jeju, and Seogwipo. One of the changes will require companies to label plastic bottles that are easy to remove. Legislative changes will also bring system reforms to make recycling more convenient. Plastic bottles with

easy-to-tear labels are produced in Japan. Designed to protect the environment from plastic pollution, it promotes the growth and innovation of industry and human life through comprehensive transformation: the production, use, and recycling of plastic bottles. PET bottles must be colorless and unlabeled to be completely recycled. It is only possible to crush transparent plastic bottles without labels into thin plastic flakes. These materials can be utilized to create new plastic items.

Plastics are an inextricable aspect of human life, particularly in countries experiencing rapid economic growth. Drinking water bottles and beverage bottles are two of the most common plastic applications in everyday life. Plastic bottles must be separated according to recyclable and non-recyclable to improve plastic bottle waste management. Recycling is the process of rebirth; plastic bottles that have been discarded are recycled into high-quality consumer goods. Recycled clear plastic bottles have been resurrected as garments, eco-friendly purses, and cosmetic bottles, among other high-quality items. Previously, all discarded plastic bottles used to make garments and other products in South Korea were imported from abroad. Only 10% of the old plastic bottles collected in the community were recycled into high-quality consumer goods. Another point to consider is that the production of plastic emits a substantial quantity of greenhouse gases, which contributes to global warming. Because recycling reduces crude oil and energy consumption, greenhouse gas emissions, such as carbon dioxide, also decrease significantly. Transparent plastic bottles are mainly used to make fiber materials for clothing, with polar fleece, a polyester material that has lately gained popularity, being a notable example. However, the foreign matter in waste bottles collected in South Korea throughout the disposal and composing procedure raises concerns about their suitability for recycling. According to the application requirements, the sorting equipment only needs to pick transparent plastic bottles in a sorting process. So correct bottle classification is crucial in the sorting system based on machine vision.

This paper proposes a GAN-based model, modified lightweight-GAN, to generate synthetic images using a small dataset containing real plastic bottle images. The main contribution is as follows:

- A new technique that enhances the imbalanced data problem using image data augmentation is proposed based on a GAN-based framework, named modified lightweight-GAN, that can generate high-quality images using a few original images.
- We propose a weighted average ensemble transfer learning-based method, IncepX-Ensemble, to classify six types of plastic bottle images.
- We construct a computationally efficient model and demonstrate its resilience based on the two presented strategies.

2. Related Works

Deep learning with a small training dataset leads to overfitting issues. The capacity to generalize data expansion was examined using deep neural network training data extensions. Instead of using traditional data augmentation techniques, GAN can generate more stable and realistic images.

A computer-aided machine learning-based plastic bottle classification technique was proposed by [5]. Specifically, the authors performed feature extraction for classification tasks by achieving 80% accuracy. The authors also proposed classification with the region of interest segmentation technique with PET and non-PET plastic bottle dataset with two classes and achieved 80% of accuracy [6]. Ref. [7] proposed an automated classification of plastic bottles based on SVM for recycling purposes and achieved 97.3% of accuracy based on the best computation time. A real-time application was designed for plastic bottle identification, and the proposed system achieved an accuracy of 97% [8]. Generative adversarial networks are an advanced technique for data augmentation and use semi-supervised cycleGAN to augment the training data. Hazra et al. proposed generating synthetic images for bone marrow cells using GAN and the classification approach using the transfer learning model [9]. The proposed model achieved 95% precision and 96% recall.

The authors of [10] proposed an inception-cycleGAN model that will classify COVID-19 X-ray images and achieved 94.2% of accuracy. An artificial intelligence-based plastic bottle color classification system was proposed by [11] and achieved 94.14% of accuracy. Wang et al. [12] proposed the recycling of used plastic bottles based on a support vector machine algorithm, and accuracy reached 94.7%. In [13], medical image classification is a famous approach; the researcher applied data augmentation using GAN and using three transfer learning models to overcome the training-time constraints. They achieved 86.45% of accuracy using the inceptionV3 model. Srivastav et al. [14] proposed an approach of generating a synthetic image using GAN to improve the diagnosis of pneumonia disease using chest X-ray image classification and achieved 94.5% accuracy. Waste management and waste classification are essential issues for the environment and human health. Recycling is one most basic forms of waste management; we need to classify the particular waste that can be recycled. There are few publicly available datasets for waste classification; for this reason, Alsabei et al. [15] proposed a model that can classify waste using pre-trained models, and for generating data, they applied the GAN approach. In [16], an intelligent system for waste sorting using a skip-connection-based model was proposed, and the novel model achieved 95% of accuracy. Pio et al. [17] hypothesized that combining a transfer learning approach with the metabolic features developed will deliver a considerable improvement in reconstruction accuracy. A new combined methodology was proposed for a higher recognition rate and robustness to enhance a low-resolution video [18]. GAN and transfer learning are used to deal with license plate image recognition in various challenging situations. Mohammed et al. [19] suggested an ensemble classifier that decreases both the space and temporal complexity of the generated ensemble members while classifying an instance by improving prediction time while maintaining significant accuracy.

3. Dataset

In our experiment, we collected plastic bottle images from the industry in South Korea. However, it is not a publicly available dataset. We intend to build models that correctly classify plastic bottle images before deploying them into a plastic bottle recycling machine. The precise detection and identification of plastic bottles is the most significant challenge when designing a recycling machine in preventing fraud. It depends on precision and cost.

There are few publicly plastic bottle datasets available. Trashnet [20] is a dataset used for trash classification that has plastic bottle images in it. Each image in the PET bottle dataset contains only one object, a plastic bottle, and a plain background. The human eye more easily perceives this but not by a recycling machine. There are no other objects in the image that could provide additional information.

Our dataset, named the PET-bottle dataset, has six classes, having a total number of 1667 plastic bottle images. We divided the plastic bottle images according to the design and bottle specification; we uniquely named three classes, Bottle_ShapeA, Bottle_ShapeB, and Bottle_ShapeC, and the other three classes are called Masinda, Pepsi, and Samdasoo, respectively. Plastic bottles which do not have a label on them but have black caps are named Bottle_ShapeA. Plastic bottles with a design on the body and a white cap but without a label are named Bottle_ShapeB. Plastic bottles that do not have any design or label on them but have a red cap are designated as Bottle_ShapeC. Masinda is a drinking water bottle company whose class depicts a company label and sky-colored cap. Pepsi is a well-known soft drink manufacturing company whose class represents a label with a company logo and black cap. Jeju Samdasoo is a mineral water brand developed by the Jeju Province Development Corporation; this plastic bottle image depicts a label with a company logo and white cap. Details of the original dataset are given in Table 1. The Sl number represents the numerical value for six classes, from 0 to 5; the class name depicts all the six classes we have used for our experiment. The images per class section describes the images containing each class.

It is noticeable that the dataset is small, and classes are primarily imbalanced in the original dataset, with most data labeled as the Samdasoo class. Training a deep neural

network to categorize the data into six categories will over-fit the data with this unbalanced dataset.

Table 1. Detailed specification of original dataset.

Sl Number	Class Name	Images per Class
0	Bottle_ShapeA	169
1	Bottle_ShapeB	238
2	Bottle_ShapeC	41
3	Masinda	249
4	Pepsi	339
5	Samdasoo	631
	Total	1667

4. Methodology

The proposed method is discussed in this section. Figure 1 depicts the proposed method's block diagram. Our proposed method can be divided into five parts. The first block (a) shows the overview of the original dataset with the class label. In the second block, (b) synthetic images are generated using a modified lightweight-GAN model for data augmentation. The third block (c) is traditional data augmentation based on basic image manipulation techniques. In the fourth block, the (d) pre-trained ImageNet model is fine-tuned on our dataset for plastic bottle classification. In the last part, (e) is the evaluation metrics for classification. A detailed explanation is given in the following subsections.
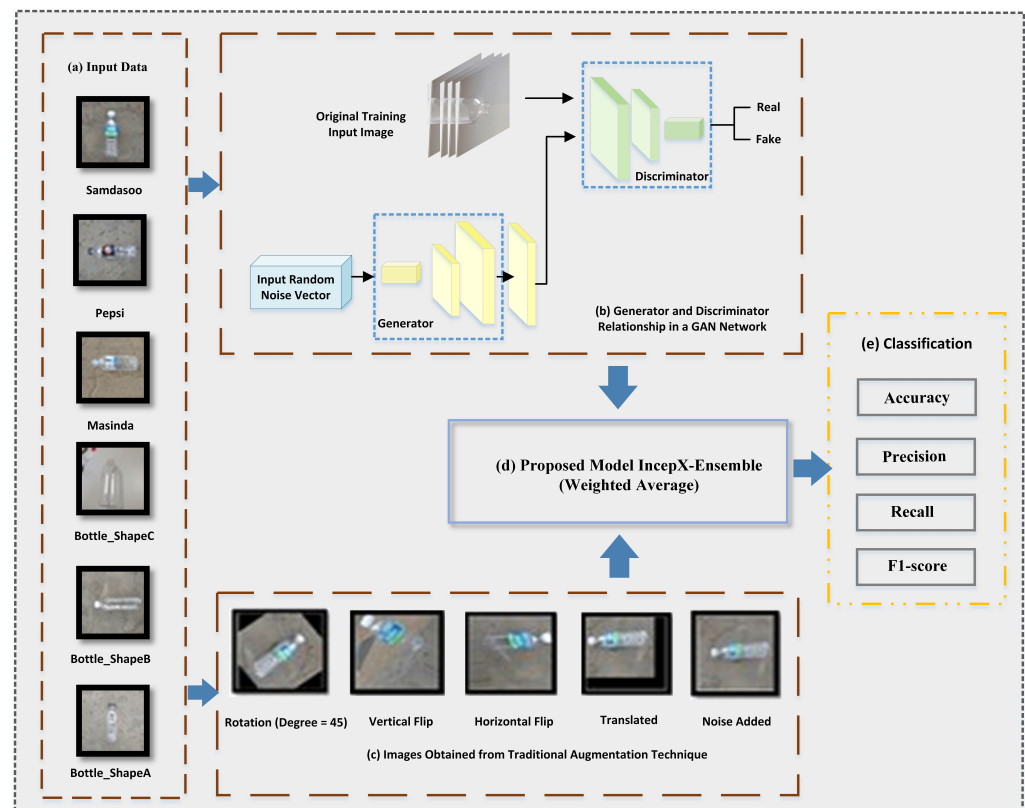
Figure 1. Workflow of the proposed framework. (**a**) shows the overview of the original dataset with the class label; (**b**) synthetic images are generated using a modified lightweight-GAN model for data augmentation; (**c**) is traditional data augmentation based on basic image manipulation techniques; (**d**) pre-trained ImageNet model is fine-tuned on our dataset for plastic bottle classification; (**e**) is the evaluation metrics for classification.

4.1. Original Dataset Description

Our dataset contains 1667 images of plastic bottles, which are segmented into six classes. The PET bottle dataset is divided into six types according to the bottle specification details.

4.2. Synthetic Image Generation Using Modified Lightweight-GAN Model

Recently, researchers have focused on combining GANs with other models or techniques that allow for superior data reconstruction. We improvised a new approach to our model. We used convolution layers compatible with high-resolution images for both G and D. The basic GAN architecture for the generator and discriminator are graphically depicted in Figure 2. The model structure of G and D and a description of the component layers are shown in Figures 3 and 4.

4.2.1. Generative Adversarial Networks

The generative adversarial network (GAN) was developed by Goodfellow et al. in 2014 [21]. This intriguing invention has been gaining interest in various machine learning fields. GAN consists of two interacting neural networks. It is a generator (*G*) and a discriminator (*D*). The generator network is trained to map points in the latent space to generate new data instances. The discriminator network is trained to distinguish between the actual and plausible images produced by the generator network. Eventually, the generator generates images that resemble actual training samples. The generator is updated based on the discriminator's predictions to have better images at the training time. The discriminator increases its ability to distinguish between actual and fake images. The difference between real and counterfeit labels determines the discriminator loss. The label specifies whether the image is artificial or natural. The general diagram of GAN is shown in Figure 2.
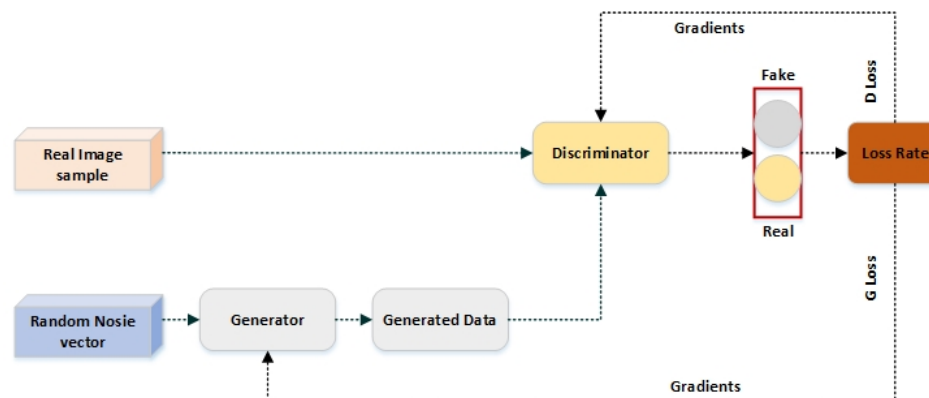

Figure 2. Generative adversarial networks architecture.

The main objective of GAN theory can be painted as a two-player min–max game which can be defined by,

$$\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x \sim P_d(x)}[log D(x)] + \mathbb{E}_{rnv \sim P_{rnv}(rnv)}[log(1 - D(G(rnv)))] \quad (1)$$

The discriminator and the generator are involved in a min–max game with the value function $V(D, G)$. The discriminator is trying to minimize its reward $V(D, G)$, and the generator is attempting to reduce the discriminator's reward or, in other words, maximize its loss.

The generator always tries to minimize the following loss function; on the other hand, the discriminator always maximizes it. In GAN, the generator receives the original input data x, adds random noise variable $P_{rnv}(rnv)$ and generates samples $G(rnv)$. $D(x)$ is the discriminator's estimate of the probability that real instance x is real over the data distribution P_d. $D(G(rnv))$ is the discriminator's estimate of the probability that a fake

instance is real. The generator tries to create almost perfect images to fool the discriminator. In contrast, the discriminator tries to improve the performance by distinguishing real and fake samples until the time that the samples generated from the generator cannot be distinguished from real data samples.

4.2.2. Generator Network

The generator needs to be impending with the deeper network to generate good synthesized images to orchestrate with high images. A deeper network means more of a convolution layer and more training time for up-sampling. Considering GPU for the training, we first fed the original image data and resized it into $128 \times 128 \times 3$. The image was scaled to $[-1, 1]$ pixel values to match the generator. It was issued because it uses the tanh activation function. The generator network inputs a 100×1 noise vector and generates fake samples. We used four convolution layers with ReLU activation to create high-quality synthesized images. Figure 3 illustrates the generator model architecture.
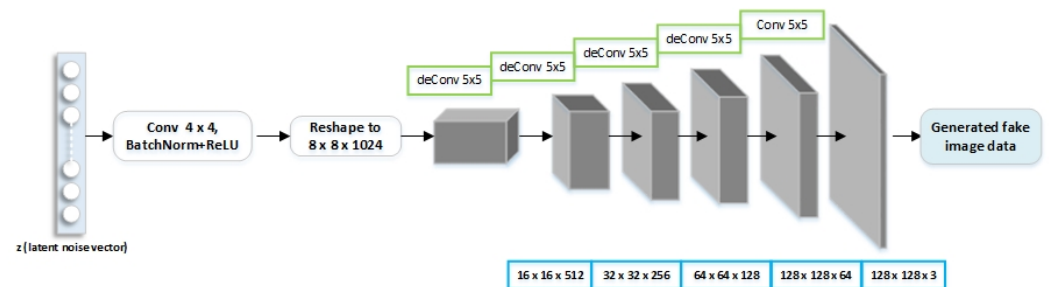

Figure 3. The architecture of the generator.

4.2.3. Discriminator Network

Following the assumption that the encoder and discriminator network information overlaps, we partially amalgamated the encoder into the discriminator [22]. The main objective of the encoder is to learn the representation feature, whereas the discriminator aims to discover the discriminating feature.

$$\mathcal{L}_{\text{recons}}^{pixel} = \mathbb{E}_{q \sim D_{encoder}(x), x \sim I_{real}}[||\kappa(q) - \tau(x)||] \tag{2}$$

where the discriminator's feature map is q, the κ function processes q, and the decoder's function τ reflects processing on sample x from real images I_{real}.

Figure 4 illustrates the discriminator model architecture. Firstly, we resize the original image to produce the I part. Then, the main part of our discriminator acts as an encoder to extract a good image feature map, and the decoder can produce a good reconstruction I'. The decoder consists of four convolutional layers to create the image 128×128. Finally, the discriminator and decoder are trained together to minimize the reconstruction loss by matching the part I' to I. The auto-encoding technique used is a common strategy for self-supervised learning that has been shown to improve model robustness and generalization capabilities [23–25].
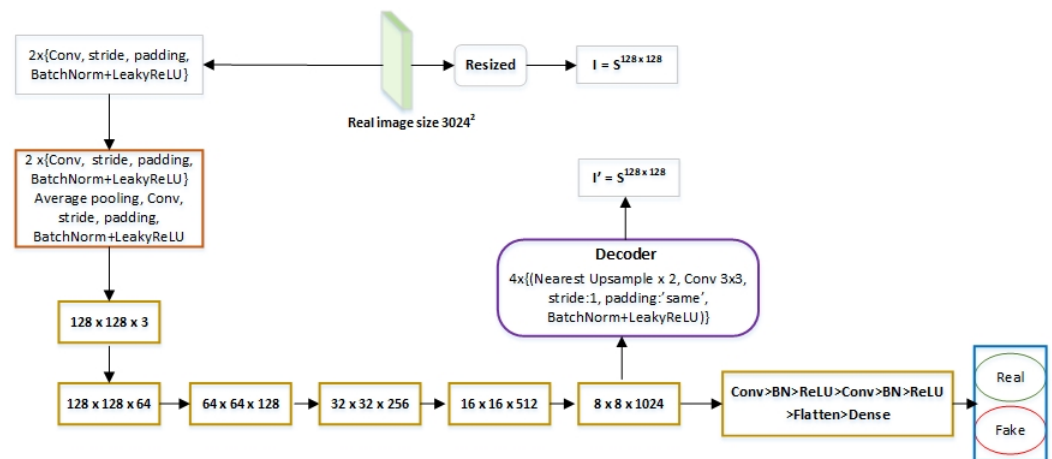
Figure 4. The architecture of the discriminator.

Recently, generative models have focused on combining new strategies with the GAN model. In many approaches, the authors combined GAN and VAE to generate a good image [22]. On the other hand, our proposed model is a pure GAN with a significantly more simple generator and discriminator and an auto-encoding function. The auto-encoding training is exclusively used for discriminator regularization and does not include the generator [26].

Here, a hinge adversarial loss for GAN is suggested, incorporating SVM margins and considering actual and fake samples falling within the margins while calculating the loss. Artificial samples outside of the boundaries that partially incorporate false local patterns are ignored in the generator training stage [27,28].

$$\mathcal{L}_{\mathrm{D}} = -\mathbb{E}_{x \sim I_{real}}[min(0, -1 + D(x))] - \mathbb{E}_{z \sim P(z)}[min(0, -1 - D(G(z)))] + \mathcal{L}_{recons}^{pixel} \quad (3)$$

$$\mathcal{L}_{\mathrm{G}} = -\mathbb{E}_{z \sim P(z)}[D(G(z))] \quad (4)$$

4.3. Traditional Data Augmentation Techniques

In this section, we describe traditional data augmentation based on basic image manipulation techniques [29]. Additionally, consider issues with limited datasets and how imbalances and data expansion can be helpful for oversampling solutions [30]. Class imbalance describes the dataset as a biased ratio of the majority to a sample of the minority.

- Flipping :
  There are two types of flipping used for image transformation; horizontal flipping is more common than vertical flipping. This augmentation is one of the simplest to employ and has shown to be effective on various datasets.

$$\begin{bmatrix} p' \\ q' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \quad (5)$$

$$p' = -p, q' = q \quad (6)$$

Horizontal flipping formulas are depicted in Equations (5) and (6).

$$\begin{bmatrix} p' \\ q' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \quad (7)$$

$$p' = p, q' = -q \quad (8)$$

Vertical flipping formulas are depicted in Equations (7) and (8).

- Rotation :
  The image is rotated right or left on an axis between [0–360] degree for rotation augmentations. The rotation degree parameter significantly impacts the safety of rotation augmentations. Outside of the rotating area, pixels are be filled with 0, and the formula of rotation is given in Equation (9).

$$R = \begin{bmatrix} cos(q) & sin(q) & 0 \\ -sin(q) & cos(q) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

  where q specifies the angle of rotation.
- Translation :
  To avoid data-position bias, shifting the image left, right, up, or down is a valuable adjustment, so the neural network looks everywhere in the image to capture it. The original image is translated into the [0–255] value range.

$$t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix} \tag{10}$$

  where in Equation (10), t_x specifies the displacement along the x axis, and t_y specifies the displacement along the y axis.
- Noise added :
  Noise is an exciting augmentation technique; noise injection injects a matrix of random values usually drawn from a Gaussian distribution. Stochastic data expansion is applied when the neural network sees the same image, which is slightly different. This difference can be seen as adding noise to the data sample and letting the neural network learn generalized features rather than overfitting the dataset.

4.4. Transfer Learning

Transfer learning techniques are used to improve the performance of machine learning algorithms using labeled data. TL efforts learn and apply one or more source tasks to enhance learning in related fields. It has been studied as a machine learning process to solve problems. TL includes pre-training models that have already been trained on large datasets and models that have been retrained at several levels of the model on a small training set. The initial layer of the pre-training network will be changed if necessary. You can use the final layer of the model's fine-tuning parameters to learn the capabilities of the new dataset [31]. According to the new task, models that have already been trained will be retrained with a smaller new dataset, and the model weights will be modified. Newly developed neural networks parameters are not built from scratch. The DL algorithm can achieve higher functionality or performance for many problems, but they need a lot of data for training time.

As a result, it can be helpful to reuse pre-trained models for similar tasks. We used two pre-trained models named inceptionV3 and Xception. The PET bottle dataset is used to fine-tune the models once they have been pre-trained with the ImageNet dataset [32]. The most common method for fine-tuning is to delete the last completely connected layer of pre-trained CNN models and replace it with a new fully connected layer (the same size as the number of classes in our dataset). Our PET bottle dataset contains six categories. Finally, the suggested method meets the goal of providing excellent classification results with a small dataset.

4.4.1. InceptionV3 and Xception

The pre-trained network models InceptionV3 and Xception were trained on millions of images from the ImageNet dataset. The InceptionV3 [33] and Xception [34] networks include 48 and 71 layers, respectively, and require a $299 \times 299 \times 3$-pixel input image. The

structure of the InceptionV3 and Xception are shown in Figures 5 and 6. While Inception considers typical congestion and yield issues, efficient results can be obtained by using asymmetric filters and bottlenecks and replacing large filters with smaller ones. Xception is simpler and more efficient. Using cross-channel and spatial correlations independently, Xception provides more specific and efficient outcomes. For the Xception model, depth-wise separable convolution is also proposed, as well as the use of cardinality to develop better abstractions.
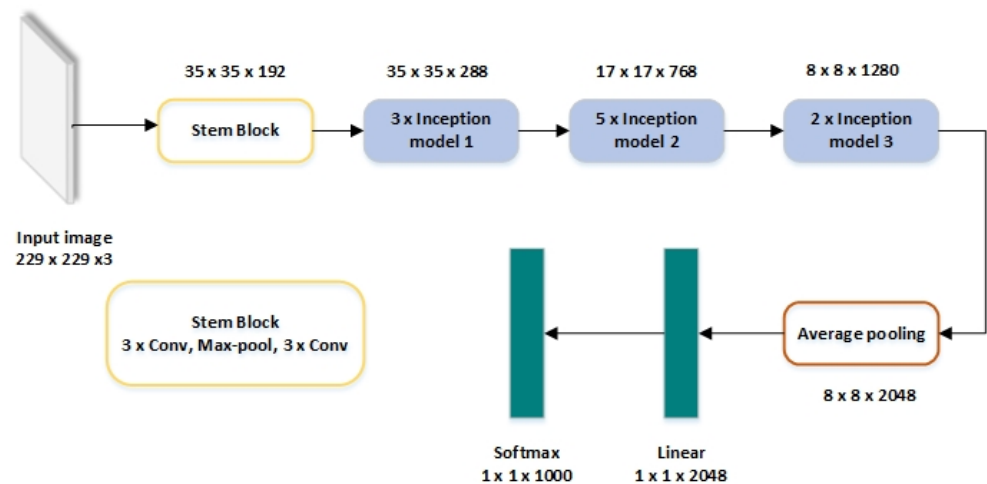

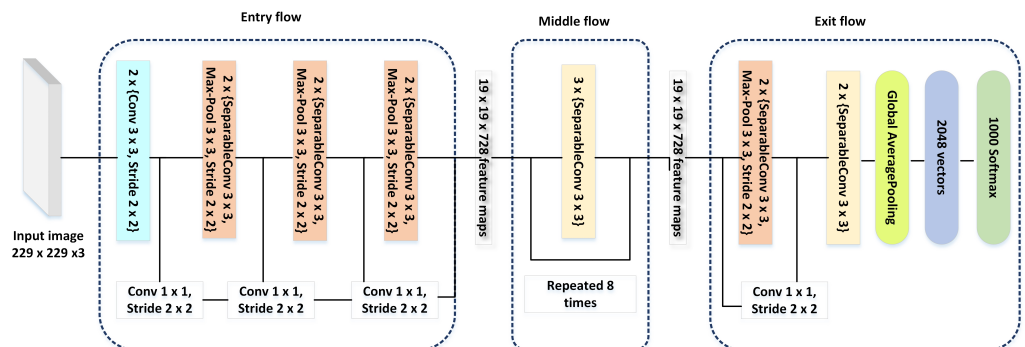**Figure 5.** InceptionV3 model architecture.

Figure 6. Xception model architecture.

4.4.2. Ensemble Learning

Ensemble learning is a way of combining multiple models to benefit in terms of computation and performance. The results of an ensemble of deep neural networks are always superior to those of a single model. The average ensemble learning was used in this study, with the same weights allocated to each model.

$$P = \frac{\sum M_i}{N} \tag{11}$$

where, in Equation (11), M_i is the probability of model i, and N is the total number of models.

DL models have different architectures and complexity; they do not provide the same result. Therefore, assigning more weights to the model performing better is convenient. By this, the maximum output can be extracted from any model. The challenge is to find the correct combination of model weights. We used the grid search technique to solve this challenge, as shown in Figure 7. A total of 1000 weight combinations were used. The search procedure continues until all varieties have been checked. The approach finally provided us with the ideal weight combination for the maximum of our given evaluation metric.
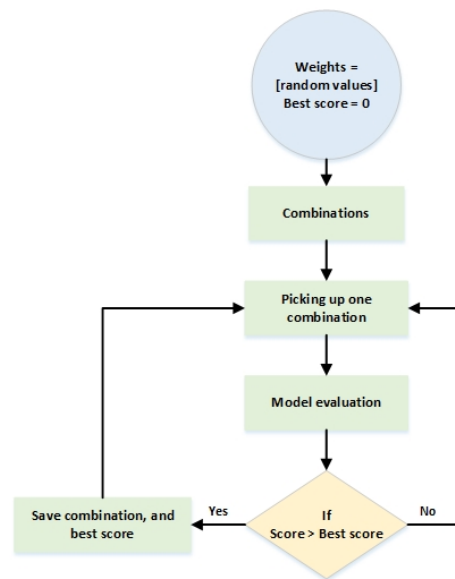
Figure 7. Grid search method for finding the weights.

4.5. Evaluation Metrics

The performance of our model was evaluated, using accuracy, precision, and recall, and the F1-score based on the confusion matrix; it includes four indicators, true positive (TP), false positive (FP), false negative (FN), and true negative (TN).

Accuracy is calculated by dividing the number of true positives and true negatives by the total number of instances. Precision is calculated with actual positive classes from the total predicted classes. The recall is derived by dividing the real positive values by the actual positive values. The F1-score is simply the average of precision and recall. Equations (12)–(15) show the accuracy, precision, recall, and F1-score calculations.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{12}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{13}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{14}$$

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \tag{15}$$

5. Results

5.1. Experimental Setup

In this study, the first part of the experiments, the modified lightweight-GAN model was trained in 500 epochs and generated synthetic images of PET bottles for each of the six categories. The weights of the generator and discriminator models were updated after each epoch to produce a composite image as close as possible to the actual image. After network training, the PET bottle dataset has 4200 images, including original and synthetic images generated from the modified lightweight-GAN model and traditional augmentation methods. In the second series of experiments, the pre-trained Inception V3 and Xception models were trained using the original training set and a combination of the training set and the image of the generated plastic bottle. Later, we employed a weighted average ensemble to enhance the classification performance using the IncepX-Ensemble model. The samples of real plastic bottle images and synthetic images generated by the modified lightweight-GAN model are shown in Figure 8. For training hyperparameter settings, we

used binary cross-entropy as the cost function, a learning rate of 0.0001, and Adam as the optimizer. We used 100 epochs and a batch size of 32 for every model.
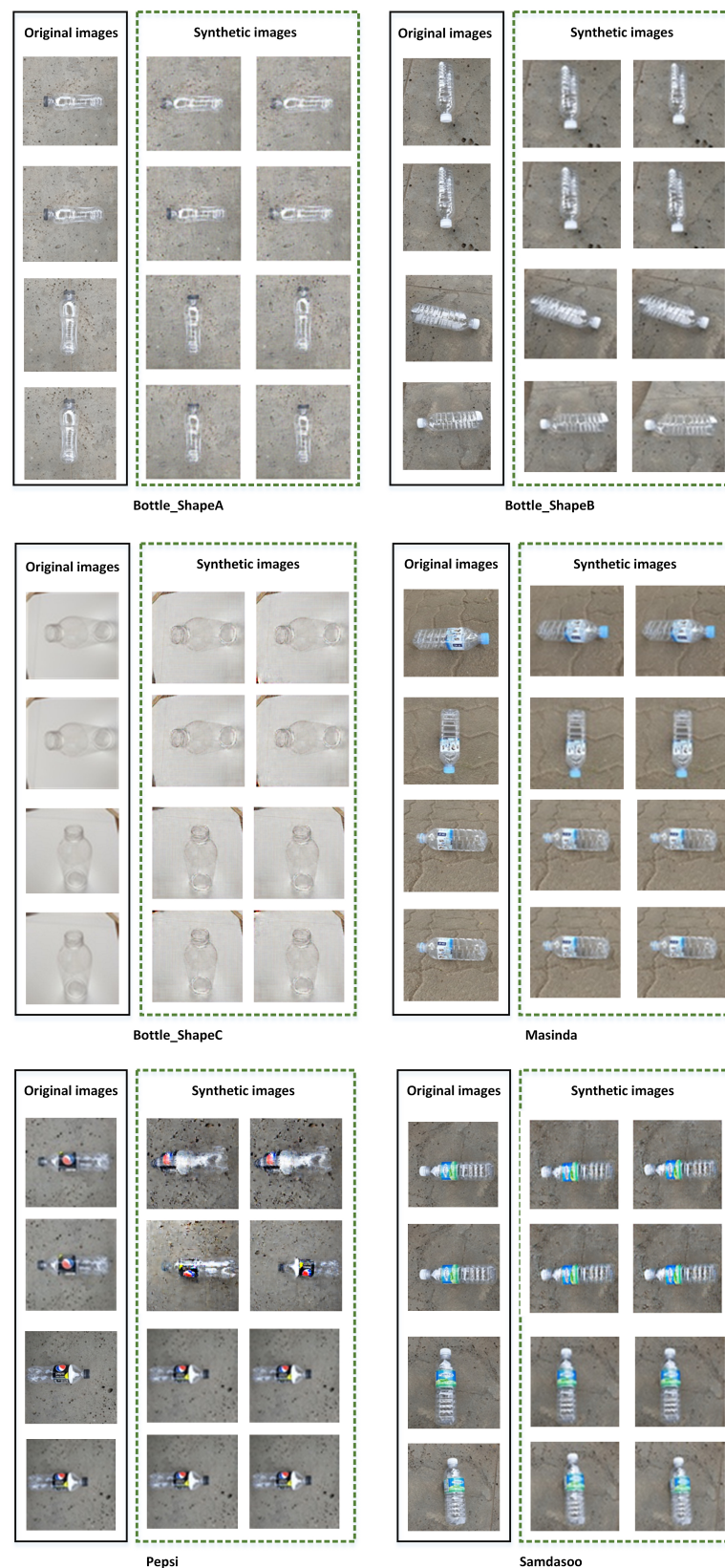

Figure 8. Original plastic bottle images and synthetic plastic bottle images generated by modified lightweight-GAN.

We divided our dataset that has 4200 images, which includes original plastic bottle images and generated images by the GAN model. Further, we split our dataset into training, validation, and testing sets for training. The training set is given to the machine learning model to analyze and learn the feature; the validation dataset is a sample of the data retained from the model training and is used to estimate the model's method while optimizing the model's hyperparameters. The test set is not used for training, and it is used to determine whether the model's hypothesis is correct. In the experiment, we first divided the dataset into 60% for training and 40% for test data. In addition, the holdout test data were split into 10% for validation (0.25% of total holdout test data) and 30% for testing (0.75% of entire holdout test data). Details of the experimental dataset are given in the Table 2.

Table 2. Details of the dataset after data augmentation using both augmentation techniques.

Sl No.	Class Name	Images per Class	Training (60%)	Validation (10%)	Testing (30%)
0	Bottl_ShapeA	700	420	70	210
1	Bottl_ShapeB	700	420	70	210
2	Bottle_ShapeC	700	420	70	210
3	Masinda	700	420	70	210
4	Pepsi	700	420	70	210
5	Samdasoo	700	420	70	210
	Total	4200	2520	420	210

5.2. Performance Metrics of GAN

We used two metrics to measure the model performance, as shown in Table 3.

- The IS is an objective metric for assessing the quality of synthetic images generated by the generative adversarial networks model. The IS was proposed by [35], and it captures the two properties of generated images: image quality, and image diversity.
- The FID is a metric that measures the overall semantic realism that compares the distance between feature vectors calculated for real and generated images. FID score was proposed by [36] to improve the performance over inception score.

Table 3. Quantitative comparison on our dataset—inception score (IS), Frechet inception distance (FID).

Sl No.	Accuracy	IS	FID
1	DCGAN	12.36	73.4
2	LSGAN	10.06	67.6
3	WGAN-GP	9.67	72.3
4	TrGAN	9.82	65.4
5	ACGAN	9.47	76.3
6	CGAN	9.89	70.0
7	Modified lightweight-GAN	9.42	64.7

5.3. Implementation Details

Specification details for performing the experiments are given below in Table 4. We used the Windows operating system with a single GPU and 32 GB of RAM. We trained our model on Tensorflow 2.6.0 version, CUDA Toolkit version 11.2 and cuDNN version 8.1.

Table 4. System components and specification.

Component	Description
Operating system	Windows 10 64 bit
Browser	Google Chrome
CPU	Intel(R) Core(TM) i5-8500K CPU @ 3.70 GHz
RAM	32 GB
Programming language	Python 3.8.5
GPU	NVIDIA GeForce RTX 2070
CUDA	CUDA Toolkit version 11.2
cuDNN	cuDNN version 8.1
Tensorflow	Tensorflow version 2.6.0
IDE	jupyter
Machine learning algorithm	Modified lightweight-GAN
Machine learning algorithm	Xception
Machine learning algorithm	InceptionV3

5.4. Classification Performance Details

In Table 5, we show how the performance of pre-trained models, such as InceptionV3 [37], Xception [38] and our ensemble model IncepX-Ensemble, may be used to determine how well classifiers can classify plastic bottle types after being trained with both original and synthetic data. The results show that the accuracy of the models is enhanced when synthetic data generated by GAN models are used to train the model. Among all the GAN models, our proposed IncepX-Ensemble model produced the best accuracy value of 99.06%.

Table 5. Comparison of IncepX-Ensemble with other existing models.

Model/Classifier	InceptionV3				Xception				IncepX-Ensemble			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Original Data	86.6	89.2	88.6	90.1	92.8	87.2	93.2	90.1	93.5	93.7	92.8	93.8
DCGAN	81.2	82.4	79.6	80.4	90.8	92.1	92.6	91.5	92.4	94.7	95.2	94.6
LSGAN	83.2	81.9	85.4	83.6	85.4	86.3	90.6	86.4	84.4	85.3	84.0	85.4
WGAN-GP	93.1	92.6	94.2	93.9	93.6	93.2	94.2	94.4	97.2	97.4	96.4	97.6
ACGAN	89.9	89.1	90.1	90.5	91.4	91.2	92.0	91.6	95.5	95.7	94.5	96.2
CGAN	97.1	98.3	96.5	97.9	98.4	97.2	98.3	97.9	97.1	98.6	98.7	98.7
Modified Lightweight-GAN	98.8	98.2	99.0	98.6	98.9	97.4	98.7	98.5	99.0	99.1	99.3	99.2

Acc, Pre, Rec, and F1 refer to accuracy, precision, recall, and f1-score, respectively.

We also assessed the performance of classification models that use original data and actual and synthetic data. We employed two different combinations of augmentation procedures for the augmentation of plastic bottle images. To produce synthetic data, Augmentation-1 employs a modified lightweight-GAN. Flipping, rotation, translation, and noise addition are all used in Augmentation-2. We kept the total number of images for each example to ensure a fair comparison.

In Table 6, we show the performance of the traditional augmentation technique with transfer learning models. We also examined classification model performance utilizing original, augmented data and a synthetic image generated by our model, which produces better quality images and performs better. We can notice that in the case of noise addition, accuracy is fairly low because of overfitting.

Table 6. Accuracy, precision, recall, and F1-score of different classification using traditional augmentation methods and a combination of original with synthetic data.

Tradition Augmentation/Classifier	InceptionV3				Xception				IncepX-Ensemble			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Original Data	86.2	75.0	86.1	86.0	86.2	75.2	89.0	86.8	88.2	87.1	94.2	89.0
Flipping	87.1	83.2	91.0	86.0	88.0	91.1	79.8	84.5	87.1	88.1	93.0	89.1
Rotation	88.5	79.7	86.5	82.2	86.1	82.0	83.5	75.8	87.0	87.1	84.1	73.0
Translation	85.1	76.5	88.1	80.2	86.2	82.2	85.1	87.5	88.1	81.1	88.0	82.2
Noise Addition	75.2	72.0	77.1	75.6	75.6	76.0	77.0	77.1	75.8	75.2	77.2	76.1
Modified Lightweight-GAN	89.8	87.4	83.7	83.3	91.3	89.3	88.5	88.7	93.1	89.6	92.9	92.1

Acc, Pre, Rec, and F1 refer to accuracy, precision, recall, and f1-score, respectively.

We evaluated our IncepX-Ensemble model with the ImgaeNet dataset in Table 7. We first trained the models with the original imageNet data and tested the model with actual data. The model can be easily adapted to support fine-tuning for classification tasks. We used the dataset for 60% for training and 40% for testing, and further testing data were split into 0.75% of the total holdout test data and 0.25% validation. The performance of the classification models using synthetic data, augmented data and a mix of original and synthetic data was then determined using the same procedure. The images created by our suggested improved lightweight-GAN model are of higher quality. It performs quantitatively better than existing GAN models, as can be seen from all of the findings.

Table 7. Evaluation of our proposed model on the ImageNet dataset.

Original + Synthetic Image/Classifier	InceptionV3				Xception				IncepX-Ensemble			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Original Data	93.9	92.5	95.8	94.3	94.4	94.6	92.9	92.9	96.2	95.8	96.1	95.6
Rotation	95.6	94.7	97.9	95.6	95.9	91.1	94.9	96.2	96.9	95.3	95.6	97.1
Translation	94.6	94.9	93.0	95.4	94.5	93.9	92.6	94.9	95.2	93.8	93.2	95.7
ACGAN	95.3	87.3	91.3	92.2	95.2	87.0	91.0	94.1	95.6	94.2	93.6	94.0
WGAN-GP	95.6	95.4	96.1	96.0	96.2	95.9	89.6	95.5	96.8	95.4	96.2	96.1
CGAN	94.6	95.0	96.1	95.3	75.6	76.0	77.0	77.1	95.8	92.5	95.4	96.0
Modified Lightweight-GAN	96.2	95.2	93.7	96.3	97.6	96.3	97.5	98.2	98.9	96.6	95.9	99.1

Acc, Pre, Rec, and F1 refer to accuracy, precision, recall, and f1-score, respectively.

6. Conclusions

The aim is to develop an application-based system that automatically detects plastic bottle images. Our proposed approach is simple: to overcome the small and imbalanced dataset, we first applied a modified lightweight-GAN method to generate synthetic images of plastic bottles. Next, we developed a transfer learning-based model, IncepX-Ensemble, classifying different plastic bottle images. Therefore, we developed a new system using the transfer learning technique, and a new framework was developed by integrating with modified lightweight-GAN. Modified lightweight-GAN was used for data augmentation enhancement of the dataset, and the proposed transfer learning-based model was trained and evaluated using original and generated images. Finally, we designed a weighted average ensemble model named IncepX-Ensemble, tuning the influence of the base models using the grid search technique. However, the two transfer learning models show excellent performance, though in some cases, the two models fail to classify plastic bottles correctly. To obtain an improved performance, we used a combination of transfer learning and the weighted average technique to boost the application performance. The obtained results indicate the algorithm's efficacy with 99.06% accuracy. Future work may validate the proposed model to evaluate recycling performance using more diverse big data. We plan to use the model we developed to explore other datasets and waste management applications

in the future. We hope that this will play a positive role in plastic bottle waste management and environmental growth.

Author Contributions: Conceptualization, S.C. and D.H.; Formal analysis, S.C.; Funding acquisition, Y.-C.B.; Methodology, S.C. and D.H.; Writing—review and editing, S.C.; Investigation, Y.-C.B.; Resources, Y.-C.B.; Project administration, Y.-C.B. and Y.-W.K.; Supervision, Y.-C.B. and Y.-W.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DL	Deep Learning
GAN	Generative Adversarial Networks
CNN	Convolutional Neural Network
TL	Transfer Learning
VAE	Variational Autoencoders
PET	Polyethylene Terephthalate
IS	Inception Score
FID	Frechet Inception Distance
DCGAN	Deep Convolutional GAN
LSGAN	Least Squares GAN
WGAN-GP	Wasserstein GAN-Gradient Penalty
ACGAN	Auxiliary Classifier GAN
CGAN	Conditional GAN

References

1. Huth-Fehre, T.; Feldhoff, R.; Kowol, F.; Freitag, H.; Kuttler, S.; Lohwasser, B.; Oleimeulen, M. Remote sensor systems for the automated identification of plastics. *J. Near Infrared Spectrosc.* **1998**, *6*, A7–A11. [CrossRef]
2. Zhang, H.; Wen, Z.G. The consumption and recycling collection system of PET bottles: A case study of Beijing, China. *Waste Manag.* **2014**, *34*, 987–998. [CrossRef] [PubMed]
3. Vo, A.H.; Vo, M.T.; Le, T. A novel framework for trash classification using deep transfer learning. *IEEE Access* **2019**, *7*, 178631–178639. [CrossRef]
4. Hammaad, S. 7.25 Million AED is the Cost of Waste Recycling. *Al-Bayan Newspaper*, 11 March 2005.
5. Ramli, S.; Mustafa, M.M.; Hussain, A.; Wahab, D.A. Histogram of intensity feature extraction for automatic plastic bottle recycling system using machine vision. *Am. J. Environ. Sci.* **2008**, *4*, 583. [CrossRef]
6. Ramli, S.; Mustafa, M.M.; Hussain, A.; Wahab, D.A. Automatic detection of 'rois' for plastic bottle classification. In Proceedings of the 2007 5th Student Conference on Research and Development, Selangor, Malaysia, 11–12 December 2007; pp. 1–5.
7. Shahbudin, S.; Hussain, A.; Wahab, D.A.; Marzuki, M.; Ramli, S. Support vector machines for automated classification of plastic bottles. In Proceedings of the 6th International Colloquium on Signal Processing and Its Applications (CSPA), Melaka, Malaysia, 21–23 May 2010; pp. 1–5.
8. Scavino, E.; Wahab, D.A.; Hussain, A.; Basri, H.; Mustafa, M.M. Application of automated image analysis to the identification and extraction of recyclable plastic bottles. *J. Zhejiang Univ.-Sci. A* **2009**, *10*, 794–799. [CrossRef]
9. Hazra, D.; Byun, Y.C.; Kim, W.J.; Kang, C.U. Synthesis of Microscopic Cell Images Obtained from Bone Marrow Aspirate Smears through Generative Adversarial Networks. *Biology* **2022**, *11*, 276. [CrossRef] [PubMed]
10. Bargshady, G.; Zhou, X.; Barua, P.D.; Gururajan, R.; Li, Y.; Acharya, U.R. Application of CycleGAN and transfer learning techniques for automated detection of COVID-19 using X-ray images. *Pattern Recognit. Lett.* **2022**, *153*, 67–74. [CrossRef] [PubMed]
11. Tachwali, Y.; Al-Assaf, Y.; Al-Ali, A. Automatic multistage classification system for plastic bottles recycling. *Resour. Conserv. Recycl.* **2007**, *52*, 266–285. [CrossRef]

12. Wang, Z.; Peng, B.; Huang, Y.; Sun, G. Classification for plastic bottles recycling based on image recognition. *Waste Manag.* **2019**, *88*, 170–181. [CrossRef] [PubMed]

13. Zulkifley, M.A.; Mustafa, M.M.; Hussain, A. Probabilistic white strip approach to plastic bottle sorting system. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 3162–3166.

14. Srivastav, D.; Bajpai, A.; Srivastava, P. Improved classification for pneumonia detection using transfer learning with gan based synthetic image augmentation. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 433–437.

15. Alsabei, A.; Alsayed, A.; Alzahrani, M.; Al-Shareef, S. Waste Classification by Fine-Tuning Pre-trained CNN and GAN. *Int. J. Comput. Sci. Netw. Secur.* **2021**, *21*, 65–70.

16. Bircanoğlu, C.; Atay, M.; Beşer, F.; Genç, Ö.; Kızrak, M.A. RecycleNet: Intelligent waste sorting using deep neural networks. In Proceedings of the 2018 Innovations in Intelligent Systems and Applications (INISTA), Thessaloniki, Greece, 3–5 July 2018; pp. 1–7.

17. Pio, G.; Mignone, P.; Magazzù, G.; Zampieri, G.; Ceci, M.; Angione, C. Integrating genome-scale metabolic modelling and transfer learning for human gene regulatory network reconstruction. *Bioinformatics* **2022**, *38*, 487–493. [CrossRef] [PubMed]

18. Du, X. Complex environment image recognition algorithm based on GANs and transfer learning. *Neural Comput. Appl.* **2020**, *32*, 16401–16412. [CrossRef]

19. Mohammed, A.M.; Onieva, E.; Woźniak, M. Selective ensemble of classifiers trained on selective samples. *Neurocomputing* **2022**, *482*, 197–211. [CrossRef]

20. Yang, M.; Thung, G. Classification of trash for recyclability status. *CS229 Proj. Rep.* **2016**, *2016*, 3.

21. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27.

22. Munjal, P.; Paul, A.; Krishnan, N.C. Implicit discriminator in variational autoencoder. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.

23. Hendrycks, D.; Mazeika, M.; Kadavath, S.; Song, D. Using self-supervised learning can improve model robustness and uncertainty. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.

24. Jing, L.; Tian, Y. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 4037–4058. [CrossRef] [PubMed]

25. Goyal, P.; Mahajan, D.; Gupta, A.; Misra, I. Scaling and benchmarking self-supervised visual representation learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 6391–6400.

26. Liu, B.; Zhu, Y.; Song, K.; Elgammal, A. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.

27. Lim, J.H.; Ye, J.C. Geometric gan. *arXiv* **2017**, arXiv:1705.02894.

28. Kim, S.; Lee, S. Spatially Decomposed Hinge Adversarial Loss by Local Gradient Amplifier. In Proceedings of the ICLR 2021 Conference, Vienna, Austria, 4 May 2020.

29. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]

30. Hao, R.; Namdar, K.; Liu, L.; Haider, M.A.; Khalvati, F. A comprehensive study of data augmentation strategies for prostate cancer detection in diffusion-weighted MRI using convolutional neural networks. *J. Digit. Imaging* **2021**, *34*, 862–876. [CrossRef] [PubMed]

31. Kamishima, T.; Hamasaki, M.; Akaho, S. TrBagg: A simple transfer learning method and its application to personalization in collaborative tagging. In Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, Miami Beach, FL, USA, 6–9 December 2009; pp. 219–228.

32. ImageNet Dataset. 2016. Available online: https://image-net.org/ (accessed on 12 July 2021).

33. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

34. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.

35. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training gans. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29.

36. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.

37. Xia, X.; Xu, C.; Nan, B. Inception-v3 for flower classification. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; pp. 783–787.

38. Wu, X.; Liu, R.; Yang, H.; Chen, Z. An xception based convolutional neural network for scene image classification with transfer learning. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 18–20 December 2020; pp. 262–267.