

## Article

# Generative Adversarial Network Models for Augmenting Digit and Character Datasets Embedded in Standard Markings on Ship Bodies

Abdulkabir Abdulraheem, Jamiu T. Suleiman  and Im Y. Jung \* 

School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, Republic of Korea; aaoabdul@gmail.com (A.A.); jamiu.suleiman111@gmail.com (J.T.S.)

\* Correspondence: iyjung@ee.knu.ac.kr

**Abstract:** Accurate recognition of characters imprinted on ship bodies is essential for ensuring operational efficiency, safety, and security in the maritime industry. However, the limited availability of datasets of specialized digits and characters poses a challenge. To overcome this challenge, we propose a generative adversarial network (GAN) model for augmenting the limited dataset of special digits and characters in ship markings. We evaluated the performance of various GAN models, and the Wasserstein GAN with Gradient Penalty (WGAN-GP) and Wasserstein GAN with divergence (WGANDIV) models demonstrated exceptional performance in generating high-quality synthetic images that closely resemble the original imprinted characters required for augmenting the limited datasets. And the evaluation metric, Fréchet inception distance, further validated the outstanding performance of the WGAN-GP and WGANDIV models, establishing them as optimal choices for dataset augmentation to enhance the accuracy and reliability of recognition systems.

**Keywords:** data augmentation; generative adversarial networks; Fréchet inception distance; ship-marking characters and digits



**Citation:** Abdulraheem, A.; Suleiman, J.T.; Jung, I.Y. Generative Adversarial Network Models for Augmenting Digit and Character Datasets Embedded in Standard Markings on Ship Bodies. *Electronics* **2023**, *12*, 3668. <https://doi.org/10.3390/electronics12173668>

Academic Editors: Galina Ilieva and George A. Tsihrintzis

Received: 24 July 2023

Revised: 28 August 2023

Accepted: 28 August 2023

Published: 30 August 2023

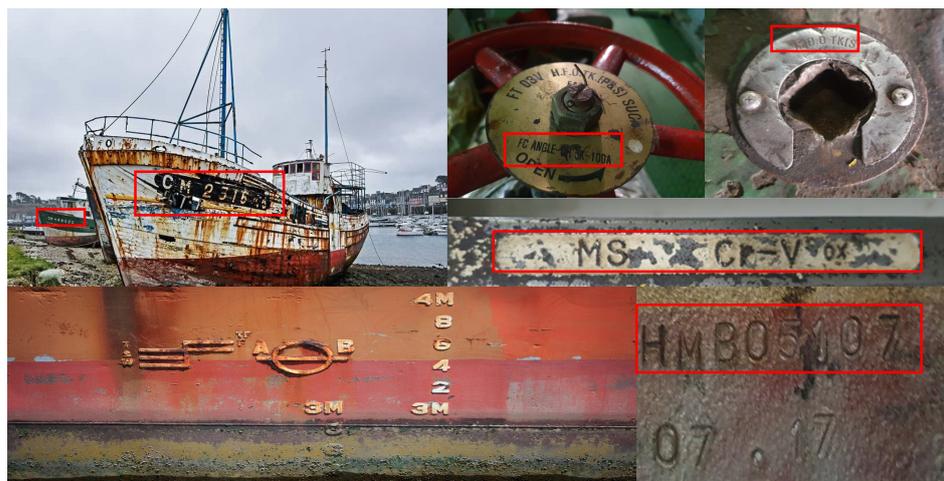


**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The recognition of characters and digits imprinted on ship bodies is of significant importance owing to their distinctive features and role in conveying crucial information. Ship markings, governed by standardized regulations, serve as identifiers offering crucial operational details [1–4]. Automatic recognition of these characters is essential for various reasons. It facilitates efficient ship documentation and tracking, enhancing maritime safety and security. Moreover, it aids in identifying vessels involved in incidents or illegal activities, supporting investigations and law enforcement. In the unfortunate event of accidents, accurate character identification provides insights for investigations and reconstructions. Imprinted character images from damaged components reveal origins and potential contributing factors. Precision in identification is vital for precise tracking, maintenance, and replacement of ship parts.

Figure 1 illustrates old ship markings, representative of our dataset sources. The dataset comprises cropped images of degraded numbers and letters from ship imprints. These real-world scenarios require recognition models resilient to degraded imprints. Unlike larger and well-maintained ships that often repaint their identification markings, older or smaller vessels face corrosion and fading, posing challenges to recognition systems. Seawater and environmental factors worsen the degradation.

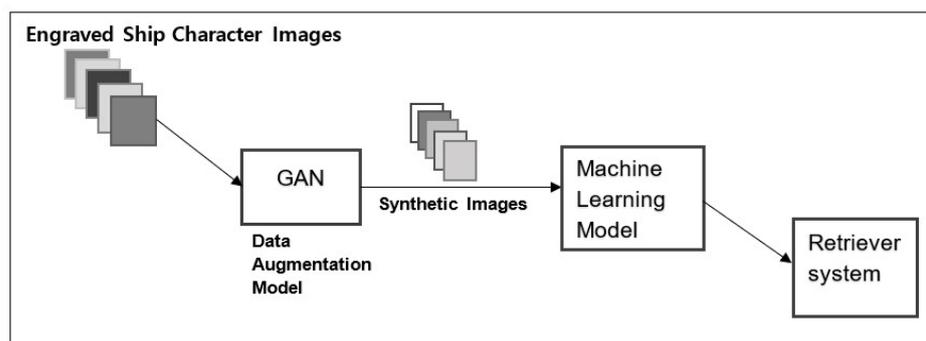


**Figure 1.** Old ship markings from sources similar to those of our dataset.

In our work, we adopt data augmentation, utilizing cutting-edge generative adversarial networks (GANs), as a promising approach, to enhance the automatic identification of imprinted characters on ships. By specifically employing GANs for data augmentation, we present a pragmatic solution tailored to scenarios where traditional augmentation methods might fall short. This is especially relevant when machine learning-based algorithms are utilized for recognition tasks. Our study thus bridges the gap between data scarcity and machine learning, offering a new perspective that can have far-reaching implications. Additionally, the significance of our work extends beyond the maritime sector. Our proposed methodology, which revolves around selecting the most suitable GAN model for a specific dataset and evaluating its performance for data augmentation, has the potential to address data scarcity challenges across diverse domains in the field of Information Technology. For instance, in domains such as medical imaging, where acquiring large datasets can be a challenge due to ethical or logistical constraints, our approach could be adapted to enhance the quality and diversity of available data [5–7].

Figure 2 depicts a whole-system-architecture chart that illustrates the information flow and operations of a system that utilizes data augmentation for enhancing the accuracy of ship-character identification and retrieval. The ship-character recognition system comprises several key components, including input data, a data augmentation sub-system, an augmented dataset, a machine learning model for classification, and a retrieval client interface. The input data comprise special alphanumeric characters found on ship components. The data augmentation module employs state-of-the-art GAN techniques to create variations of the original images. The augmented dataset comprises the synthetic images produced by the GAN. The machine learning model learns to identify the imprinted characters on the ship components using the augmented dataset. Furthermore, the retrieval system retrieves relevant ship-component information, such as part numbers, specifications, and maintenance history based on the identified characters and digits.

The subsequent sections of this paper are structured as follows: First, we review and analyze previous studies on augmentation techniques in Section 2 for extending object detection in maritime images. Then, we present our methodology in Section 3, outlining the experimental setup and procedures employed to evaluate the performance of cutting-edge GAN models on ship-marking-character and -digit datasets. Following that, we provide a comprehensive evaluation of our results, analyzing the outcomes and comparing the generated character and digit images. We also discuss the limitations encountered during the study, shedding light on potential constraints and areas for improvement. Furthermore, we present the potential direction of future research suggesting enhancements in the proposed approach. Lastly, we conclude this article by summarizing our findings and highlighting the significance of our research in advancing the augmentation of ship characters.



**Figure 2.** Ship-character image recognition process.

## 2. Related Work

Extensive research focused on achieving reliable detection and recognition of ship images utilizing different techniques has been performed. While considerable progress has been made, our work contributes to this research area by providing an extensive and diverse dataset for training recognition models. We reviewed previous studies on ship identification and data augmentation with the aim to gain valuable insights into the effectiveness of data augmentation approaches and their effect on improving ship-character identification. Some of the works mentioned here explore the use of convolutional neural networks (CNNs) and GANs in the ship application domains.

In [8], the authors addressed the important issue of ship-type identification in maritime surveillance. They highlighted the challenges in building large-scale marine-environment datasets, where data collection and security concerns limit the availability of comprehensive data. To overcome these limitations, the authors proposed a novel approach utilizing GANs for data augmentation. By augmenting a small number of real ship images, they improved fine-grained ship classification performance and demonstrated the effectiveness of augmented data in training ship classification networks. This research demonstrates the potential of augmented data for enhancing ship identification and classification for maritime surveillance. Ref. [9] proposed a data augmentation method for extending object detection datasets in maritime images. Their approach involved extracting the mask of the foreground object and combining it with a new background to generate location information and additional data. This technique aimed to enhance the learning process by incorporating diverse and high-quality data features. Further, experimental evaluation demonstrated the effectiveness of their method in improving the performance and robustness of object detection models specifically tailored to maritime imagery. Ref. [10] introduced BoxPaste, a powerful data augmentation method tailored for ship detection in Synthetic Aperture Radar (SAR) imagery. Their approach involves pasting ship objects from one SAR image onto another, thereby achieving considerable performance improvements in the SAR ship-detection dataset compared with baseline methods. They also proposed a principle for designing SAR ship detectors, emphasizing the potential benefits of lighter models. The integration of their data augmentation scheme with RetinaNet [11] and Adaptive Training Sample Selection (ATSS) [12] further demonstrates its effectiveness, resulting in impressive performance gains.

In [13], the authors proposed a modification to the Faster R-CNN object detection network to tackle the challenge of multiscale ships in SAR images. By incorporating the constant false-alarm-rate algorithm and re-evaluating low-scoring bounding boxes, the proposed method achieved improved detection performance. This work contributes to the advancement of SAR ship detection using deep learning methods and provides valuable insights for addressing the multiscale-ship-detection problem. Ref. [14] proposed a densely connected multiscale neural network based on the faster R-CNN framework for multiscale and multiscene SAR ship detection. Their method addressed the challenges in detecting small-scale ships and handling complex backgrounds in SAR images. By densely connecting feature maps and introducing a training strategy to focus on hard examples,

their approach achieved excellent performance in multiscale SAR ship detection across various scenes.

Data augmentation across multiple domains and the use of simple models trained on large datasets can be highly beneficial for the performance improvement of object detection applications. The effectiveness of this approach in improving the accuracy and robustness of object detection algorithms was demonstrated in [15–17]. By augmenting available data with various transformations, such as rotations, translations, and noise addition, the models can better generalize and exhibit improved detection capabilities across various scenarios and variations in the input data. Ref. [18] addressed the challenge of training deep learning models that require a large number of images by employing data augmentation as a preprocessing step. They resized dataset images to a uniform size of  $256 \times 256$  pixels and applied various augmentation techniques, such as right shift, image flipping, and left shift. These methods increased dataset diversity and improved the performance of the classifier. Image flipping, in particular, allowed for modifications in the pixel location of pixels, enhancing the variations within the dataset and enabling more robust model training. The authors in [19] addressed the problem of insufficient data by utilizing Conditional Wasserstein GAN-Gradient Penalty (CWGAN-GP), and DenseNet and ResNet. They employed GANs to generate underwater sonar images and expanded the dataset. GANs have gained considerable attention due to their ability to learn complex data distributions in high-dimensional spaces. By employing CWGAN-GP&DR, the authors addressed the overfitting issue and successfully expanded the dataset for improved model training. Ref. [20] introduced several data augmentation techniques for improving palimpsest character recognition using deep neural networks. Palimpsests are manuscripts with overlaid text that makes recognizing the underlying characters challenging. The authors proposed four augmentation methods—random mask overlay, random rotation, random scaling, and random noise addition. These methods were evaluated on a palimpsest dataset, and they observed that the random mask overlay method achieved the best performance, improving character recognition accuracy by up to 10%. These findings highlight the effectiveness of data augmentation in enhancing the performance of character recognition algorithms for palimpsest manuscripts.

In [21], the authors proposed a generative and discriminative model-based approach for information retrieval. Their minimax game model generates text queries that are relevant to a given document and learns to discriminate between relevant and irrelevant documents. The model achieved state-of-the-art results in various information retrieval tasks.

While different from our approach, the methods in these papers provide valuable insights into scene text detection, recognition, and segmentation. Furthermore, these studies showcase the progress made in understanding and processing textual information in complex visual environments, contributing to our understanding of character recognition and augmenting our knowledge in the domain of ship-character data augmentation. Our approach emphasizes the importance of equipping recognition models with abundant data from diverse character datasets collected from multiple ships and ship-body images. This extensive dataset encompasses various ship markings, allowing the recognition model to effectively learn and generalize across different ship types and marking variations. By leveraging this rich dataset, we aim to enhance the accuracy and robustness of ship-character recognition, enabling more effective ship identification mechanisms in real-world scenarios.

### 3. Dataset and Data Augmentation Methods

This section provides an overview of the datasets, the state-of-the-art GAN techniques, and the evaluation techniques employed in this work and offers valuable insights into the intricacies of the methodology and its relevance in the field of ship-character recognition. The presentation of datasets and the demonstration of the utilization of advanced GAN

techniques enhance the understanding of the nuances of our methodology and emphasize its importance in ship-character analysis and recognition.

### 3.1. Datasets

The dataset used in the experiments comprises ship-character images (0–9 digits and 13 letters (A, C, D, E, I, L, M, N, O, P, R, S, and T)) obtained from old or poorly maintained ships. These images were carefully selected to support ship-character identification and retrieval systems. These images exhibit various characteristics to capture the diverse ship markings found on different parts of body and engine components, as depicted in Figure 3.



**Figure 3.** Collection of character and digit samples from the source dataset.

The images in the dataset exhibit variations in size and color, but they were pre-processed to ensure consistency during training and analysis. Specifically, they were normalized to grayscale and resized to  $56 \times 56$  pixels in width and height. The images are stored in standard formats such as JPEG or PNG. The dataset encompasses various engraving styles commonly found on ships, including embossed, engraved, and painted characters, either individually or in combination. These characters represent ship identification numbers, hull markings, engine-component identifiers, and other characters relevant to ship operations. This curated dataset serves as the foundation for evaluating and comparing the performance of GAN models in generating synthetic ship-character images, thereby enabling the development of accurate and reliable ship technology applications.

### 3.2. State-of-the-Art GANs

We selected specific GAN models for comparison based on their application areas and their ability to generate high-quality images in a shorter time frame compared with photorealistic GAN models. Our selection took into account the practicality and efficiency of generating synthetic images for our research goals. Augmenting data using GANs and then using them for network training is a considerably useful method for avoiding infringement of personal information and security problems in data [8]. Our experiment was performed using the following GAN models:

- GAN [22]: GAN is a fundamental model in which a generator and discriminator are trained in an adversarial manner. The generator aims to produce synthetic samples, while the discriminator distinguishes between real and fake samples. GANs have demonstrated their ability to generate realistic data across various domains.
- Auxiliary Classifier GAN (AC-GAN) [23]: AC-GAN extends the conditional GAN framework by having the discriminator predict the class label of an image instead of receiving it as input. This approach stabilizes training, allows the generation of large, high-quality images, and promotes a latent space representation independent of the class label.

- Boundary-Seeking GAN (BGAN) [24]: BGAN focuses on learning the manifold boundary of the real data distribution by minimizing the classification error of the discriminator near the decision boundary. This encourages the generator to generate samples that lie on the data manifold, resulting in higher-quality and more realistic generated samples.
- Boundary Equilibrium GAN (BEGAN) [25]: BEGAN optimizes a lower bound of the Wasserstein distance using an autoencoder as the discriminator. It maintains equilibrium between generator and discriminator using an additional hyperparameter.
- Deep Convolutional GAN (DCGAN) [26]: DCGAN utilizes CNNs as the generator and discriminator. It introduces architectural constraints to ensure the stable training of CNN-based GANs and demonstrates competitive performance in image classification tasks.
- Wasserstein Generative Adversarial Network (WGAN) [15]: WGAN utilizes the Wasserstein distance to measure the discrepancy between real and generated data distributions. It introduces a critic network and focuses on optimizing the Wasserstein distance for stable training.
- WGAN with GP (WGAN-GP) [27]: WGAN-GP proposes a GP to enforce the Lipschitz constraint in the discriminator, replacing the weight clipping used in WGAN. This penalty improves stability, prevents issues such as mode collapse, and eliminates the need for batch normalization.
- Wasserstein divergence (WGANDIV) [28]: WGANDIV approximates Wasserstein divergence; exhibits stability in training, including progressive growing training; and has demonstrated superior quantitative and qualitative results.
- Deep Regret Analytic GAN (DRAGAN) [29]: DRAGAN applies a GP similar to WGAN-GP but with a focus on real data manifold. Even though DRAGAN is similar to WGAN-GP, it exhibits slightly less stability compared with WGAN-GP.
- Energy-based GAN (EBGAN) [30]: EBGAN models the discriminator as an energy function that assigns low energies to regions near the data manifold. It focuses on capturing regions close to the data distribution.
- FisherGAN [31]: FisherGAN introduces GAN loss based on the Fisher information matrix, maximizing Fisher information to encourage diverse and high-quality sample generation. It improves mode coverage and sample quality, enhancing the performance of GANs in generating realistic and varied data.
- InfoGAN [32]: InfoGAN extends the GAN framework by introducing an additional latent variable that captures the interpretable factors of variation in the generated data. By maximizing the mutual information between this latent variable and the generated samples, InfoGAN enables explicit control over specific attributes of the generated data. It promotes disentangled representations and targeted generation.
- Least-squares GAN (LSGAN) [33]: LSGAN addresses the vanishing gradient problem using the least-squares (L2) loss function instead of cross-entropy. It stabilizes the training process and produces visuals that closely resemble real data.
- MMGAN and Non-Saturating GAN (NSGAN) [34]: NSGAN simultaneously trains the generator (G) and discriminator (D) models. The objective is to maximize the probability of D making a mistake. NSGAN differs from MMGAN in its generator loss. Furthermore, the output of G can be interpreted as a probability.
- RELATIVISTIC GAN (REL-GAN) [35]: It introduces a relativistic discriminator that compares real and generated samples in a balanced manner by considering their relative ordering. This approach reduces bias toward either real or fake samples, resulting in improved training stability and generation quality.
- SGAN [36]: SGAN maintains statistical independence between multiple adversarial pairs, addresses limitations in representational capability, and exhibits improved stability and performance compared with standard methods. SGAN is suitable for various applications and produces a single generator. Future extensions can explore diversity between pairs and consider multiplayer game theory.

We implemented all of the GANs listed above by adapting and optimizing them for our dataset, executed them for our dataset, evaluated the results using well-established metrics, and then selected the GAN models that produced diverse and high-quality images of the imprinted characters. The selected models were used to generate additional images of imprinted characters. During the training phase, our focus was on characters (specifically, the letters A, C, D, E, I, L, M, N, O, P, R, S, and T) and digits (0–9). These specific characters and digits were chosen because of their easy availability and the convenience of collecting them from various sources. These characters and digits were grouped into classes according to each character or digit, and each class was trained for a specific number of epochs—20,000 and then 50,000 epochs. During each epoch, the GAN models processed each training example in the dataset, calculated the loss, and updated the model parameters using the chosen optimization algorithm. The number of epochs determined the frequency at which the model iteratively updated its parameters to learn from the data. An epoch was considered complete when all the training examples had been used once for parameter updates. Employing multiple epochs is a common practice for improving the performance of the model by allowing it to learn from the data multiple times. To assess the performance of the GAN models and provide a rationale for their selection, we performed visual inspections of the generated images [37–39]. We implemented each GAN model based on the original design proposed by the respective authors, with minimal hyperparameter tuning. Our objective was to identify the most suitable and efficient GAN model for our specific use case. For the training process, we trained each GAN model for 20,000 and 50,000 epochs. A relatively high number of epochs was selected to ensure adequate learning and convergence of the models. Throughout the training process, image outputs at the 50th and 100th iterations were generated to monitor the progress and visually assess the generated samples. Consistency with the original recommendations was maintained as closely as possible. However, certain GAN models had unique parameters and architectural variations. Our focus was to maintain consistency with the recommended settings and focus on evaluating the overall performance and quality of the generated images across the different GAN models.

### 3.3. Evaluation Metrics

We evaluated the results using well-established metrics, and the GAN models that exhibited both diversity and high-quality images of the imprinted characters were selected. Subsequently, the chosen models were utilized to generate additional images of imprinted characters.

Evaluating the quality and fidelity of generated images in a GAN presents unique challenges due to the absence of a universal discriminator for fair comparisons. When assessing GAN performance, two primary properties—fidelity and diversity—must be considered. Fidelity refers to the realism and visual quality of the generated images, taking into account factors such as image clarity and resemblance to real samples, whereas diversity measures the range and variety of images produced by the generator, ensuring that it captures the entire scope of the training data or desired modeling class. Evaluating fidelity involves comparing generated samples to their closest real counterparts and analyzing the overall distribution of fake versus real images. Diversity evaluations require the evaluation of the ability of the GAN model to generate diverse images rather than producing a single realistic but limited output. Striking a balance between fidelity and diversity is crucial, as a successful GAN should consistently generate high-quality images while covering a wide range of possibilities. However, accurately quantifying these properties remains a challenge, particularly without relying on memorizing the training dataset. By considering fidelity and diversity, evaluators can gain valuable insights into the performance of the GAN model and its capability to generate convincing and varied fake images.

Visual examination of samples is one of the most common and intuitive ways to evaluate GANs. However, it has several limitations, including the reviewer's biases toward the model, its configuration, and the project objectives [40]. In addition, visual examination

requires knowledge of what is realistic and what is not for the target domain, and it is limited to the number of images that can be reviewed in a reasonable time.

The evaluation of GAN models encompasses various methods, such as Fréchet inception distance (FID) [41], Inception Score (IS) [37], and precision and recall [42]. These metrics provide valuable insights into different aspects of the generated images, including their quality, diversity, and resemblance to real data.

The IS is a commonly used metric for evaluating the quality and diversity of generated images [37]. A higher score indicates better performance with low entropy in the conditional probability distribution and high entropy in the marginal probability distribution. However, the IS has several limitations. It can be easily manipulated or exploited to achieve high scores by generating one real image per classifier class, resulting in a lack of diversity. Furthermore, it solely considers the generated samples and does not compare them to real images. The proxy statistics used in the calculation may not accurately reflect real-world performance and are dependent on the tasks and capabilities of the classifier. Additionally, the IS may not provide precise results when dealing with images containing multiple objects, as it is trained on the ImageNet dataset, which focuses on single-object classification. Given that our imprinted digit dataset does not align with the ImageNet classes, the IS metric may not offer meaningful insights into the quality and diversity of the generated images [43].

The FID is commonly used to evaluate GANs by measuring the similarity between real and generated images based on their embeddings.

$$FID(x, g) = \|\mu_x - \mu_g\|_2^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}) \quad (1)$$

As shown in Equation (1),  $FID(x, g)$  for a 'multivariate' normal distribution calculates the Fréchet distance and aims for a lower value, indicating better performance [39,41,44,45].  $x$  and  $g$  are the real and fake embeddings (activation from the Inception model) assumed to be two multivariate normal distributions.  $\mu_x$  and  $\mu_g$  are the magnitudes of vectors  $x$  and  $g$ .  $\text{Tr}$  is the trace of the matrix, and  $\Sigma_x$  and  $\Sigma_g$  are the covariance matrices of the vectors. We assessed the quality, fidelity, and overall resemblance of the generated images to the original characters and digits. Additionally, we used the FID as an objective evaluation metric, which allowed us to quantitatively measure the proximity of the generated images to the real images, thereby providing further validation and justification for selecting the preferred GAN model [39,41,44].

A high precision value indicates that the discriminator correctly identifies a large proportion of the generated samples as fake, minimizing false positives. On the other hand, a high recall value indicates that the discriminator correctly identifies a large proportion of the real samples as genuine, minimizing false negatives [42]. These metrics further contributed to the evaluation of GAN performance and the assessment of the ability of discriminators to distinguish between real and generated images.

The IS primarily focuses on the diversity of the generated images and does not consider the efficiency of the generator in approximating the real image distribution. Thus, it is limited in its ability to measure the fidelity to real images. On the other hand, the FID can detect intra-class mode dropping and provides a more comprehensive evaluation metric by considering the quality and diversity of the generated samples. However, the IS and the FID have limitations in detecting overfitting. Precision and recall metrics are impractical for real images as the underlying data manifold is usually unknown, making them only suitable for evaluations on synthetic data, where the ground truth is available. Thus, although precision and recall are relevant metrics in certain scenarios, they are not widely applicable or suitable for assessing the performance of generative models on real-world image datasets [44]. Despite these limitations, the FID is widely used for evaluating generative models due to its robustness, reliability, and consistency in comparing model performance, particularly when dealing with large sample sizes [44]. Particularly, its consistency in relative model comparisons makes it a preferred choice among researchers and practitioners in the field.

We chose the GAN models that demonstrated low FID values and exhibited promising results in terms of generating high-quality images. Moreover, we considered the average training time of the models, taking into account computational efficiency and the practicality of generating images within a reasonable time frame. Owing to the combination of visual inspection and quantitative evaluation metrics, the selected GAN models not only produced visually appealing images but also met the desired efficiency and performance requirements for our research.

To identify the most suitable GAN model for our GAN augmentation task, we performed an extensive evaluation of FID scores for various GANs, including ACGAN, BGAN, BEGAN, DCGAN, DRAGAN, EBGAN, FISHERGAN, GAN, INFOGAN, LSGAN, MM-GAN, NSGAN, RELATIVISTIC GAN, SGAN, WGAN, WGAN-GP, and WGANDIV. Each GAN model was trained, and its generated images were subjected to FID analysis. The FID score, which measures the dissimilarity between the feature distributions of real and fake images, was calculated for each GAN. The lower the FID score, the closer the resemblance between the generated and real images, indicating higher image quality [45].

#### 4. Results and Evaluation

In our evaluation, we focused on generating the 13 available letters (A, C, D, E, I, L, M, N, O, P, R, S, and T) and digits (0–9). To achieve optimal performance, we carefully tuned the hyperparameters of the GAN models. The Learning Rate (LR) determines the step size at which the model updates its parameters during training. A lower LR results in more stable training but slower convergence. The batch size refers to the number of samples processed in each iteration. A larger batch size can accelerate training but may require more memory. Dropout is a regularization technique that randomly drops out a fraction of the units of the model during training to prevent overfitting and promote generalization. The number of epochs determines how many times the model will iteratively update its parameters to learn from the data. An epoch is completed when all training examples have been used once for parameter updates. We introduced random noise as input vectors for the generator across all GANs to ensure randomness and diversity. We chose a batch size of 32 to prevent overfitting, considering the size of our dataset. We extended the epoch count from 20,000 to 50,000 in all experiments. We also generated sample outputs periodically for visual assessment. We evaluated the performance of all models using four different learning rates: 0.001, 0.002, 0.0001, and 0.0002. However, we used the Adam optimizer with a learning rate of 0.0002 for both generators and discriminators for WGAN-DIV and WGAN-GP. These adjustments were made based on empirical evaluations to determine the optimal values that result in improved GAN performance. After training and evaluating various models, WGANDIV, WGAN-GP, GAN, and BGAN emerged as the top-performing models, exhibiting exceptional visual appeal in their generated outputs. Upon closer examination, WGAN-GP and WGANDIV exhibited similar characteristics and were visually comparable and better. These two GAN models demonstrated superior performance after being trained for 50,000 epochs, and the optimal output could be achieved around 12,500 epochs.

Tables 1 and 2 present the FID scores achieved by each GAN model when applied to the character and digit datasets, respectively. A lower FID score indicates higher similarity and better quality of the generated samples, reflecting the effectiveness of the GAN model in capturing the underlying characteristics of the dataset. Upon analyzing the results, it was observed that WGANDIV exhibited the lowest FID score among all the evaluated GAN models, indicating its superior image quality. This exceptional performance establishes WGANDIV as the preferred GAN model for our data augmentation task because it excels at generating highly realistic images. Additionally, WGAN-GP also exhibited commendable performance, indicating its effectiveness in producing high-fidelity images. Thus, WGANDIV and WGAN-GP were the top-performing GAN models in our evaluation.

**Table 1.** FID values for the evaluated GANs.

GAN	A	C	D	E	I	L	M	N	O	P	R	S	T
ACGAN	412.8	482.6	469.5	496.6	414.3	442.6	453.5	455.1	419.7	481.0	407.9	570.5	470.2
BGAN	299.8	298.2	242.0	229.1	311.1	296.9	272.1	294.7	312.5	340.5	268.3	287.6	297.3
BEGAN	402.6	434.3	569.3	398.9	365.9	315.4	453.2	438.6	328.1	304.8	249.2	465.0	320.9
DRAGAN	375.5	472.6	416.4	442.0	512.7	450.8	485.5	469.1	430.9	437.1	472.9	467.9	462.7
EBGAN	581.3	434.9	470.3	400.4	457.3	384.8	434.6	408.9	383.9	401.7	381.0	514.4	436.1
F-GAN	525.7	499.1	441.7	497.1	462.5	530.0	553.3	498.6	447.2	491.7	502.6	551.3	526.2
GAN	242.6	316.3	258.4	218.0	316.8	273.8	280.0	260.9	300.5	232.9	245.8	295.7	229.4
INFOGAN	479.1	481.7	443.7	493.0	529.9	515.2	522.9	484.5	438.9	449.9	492.8	505.5	527.6
LSGAN	393.5	409.8	415.1	411.6	393.8	361.1	377.9	400.8	419.6	356.8	384.4	472.1	463.4
MMGAN	455.9	486.2	419.2	408.4	538.5	495.5	526.4	464.5	466.2	410.1	486.2	488.6	512.5
NSGAN	449.1	478.7	416.0	408.9	519.7	453.6	486.6	459.4	432.4	436.9	463.1	485.8	478.1
REL-GAN	300.1	359.1	383.4	393.7	368.4	385.2	376.5	432.8	434.5	391.9	373.8	468.1	421.3
SGAN	350.4	407.9	424.3	419.5	365.6	372.8	411.3	381.8	349.6	429.4	406.2	493.2	389.0
WGAN	380.3	313.5	349.5	258.5	293.2	288.5	369.9	332.6	324.9	249.7	368.6	337.1	391.5
WGAN-GP	261.7	271.8	188.6	197.8	268.6	236.5	247.8	294.2	230.0	211.2	258.8	307.8	290.3
WGANDIV	231.6	224.8	210.2	215.8	279.1	241.0	252.1	245.5	283.3	213.4	255.0	290.4	285.4

**Table 2.** FID values for the evaluated GANs.

GAN	0	1	2	3	4	5	6	7	8	9
ACGAN	352.2	442.7	350.9	339.7	381.0	419.9	400.543	383.9	320.8	387.5
BGAN	250.4	278.6	282.0	269.1	314.9	220.4	286.6	293.3	292.9	291.5
BEGAN	312.9	379.8	423.1	367.5	365.9	368.1	404.4	463.9	318.1	346.7
DRAGAN	320.5	355.6	394.0	377.4	375.83	357.3	359.8	357.4	383.4	368.3
EBGAN	321.8	370.4	390.0	377.4	373.4	351.3	443.6	411.7	398.0	373.9
FISHERGAN	417.3	502.9	487.3	477.9	519.3	395.5	421.6	408.1	465.4	522.7
GAN	273.9	251.1	262.2	262.1	281.3	266.8	253.0	284.1	286.1	263.2
INFOGAN	301.1	363.6	393.2	412.8	405.5	406.5	349.7	390.9	412.8	386.4
LSGAN	327.6	372.2	291.4	339.0	356.4	339.4	355.3	349.4	366.8	395.7
MMGAN	438.4	339.8	351.8	349.0	366.3	390.6	322.3	344.1	368.5	412.3
NSGAN	307.5	401.4	400.2	418.8	390.7	419.8	330.4	373.6	388.6	402.7
REL-GAN	333.1	356.1	286.6	340.3	321.5	335.1	411.5	384.3	394.4	414.6
SGAN	322.6	389.7	342.3	362.1	374.5	374.3	389.7	356.5	370.7	391.5
WGAN	247.8	305.2	318.4	353.0	342.0	339.9	383.0	383.0	395.4	330.6
WGAN-GP	224.0	305.0	236.9	229.3	246.9	240.3	231.2	256.0	262.3	289.5
WGANDIV	235.6	289.5	239.9	223.2	358.9	220.7	256.3	270.6	239.5	299.7

The success of WGAN-GP and WGANDIV can be attributed to several factors. First, both models utilize Wasserstein distance as a loss function, which helps address the mode-collapse issue commonly encountered in GAN training. WGAN-GP employs the gradient penalty technique to enforce Lipschitz continuity, promoting stable training and preventing mode collapse. On the other hand, WGANDIV incorporates an additional divergence term that encourages the generator to produce more diverse samples, resulting in improved quality. To further support our evaluation, we performed a detailed visual inspection of the generated samples from the four best GAN models. Based on the original images provided in Figure 4, Figures 5–8 provide representative images showcasing the outputs from BGAN, GAN, WGAN-GP, and WGANDIV, respectively. Upon visual examination, it is evident that the samples generated by WGAN-GP and WGANDIV shown in Figures 7 and 8 exhibit superior quality in terms of capturing the intricate details and characteristics of the imprinted characters and digits. The images from these models demonstrate sharper edges, more pronounced textures, and enhanced overall fidelity compared with the other models with very-low-quality images, as shown in Figures 9–13.

During the evaluation process, we observed variations in the FID scores across various numbers and characters, regardless of the GAN models used. This discrepancy in FID

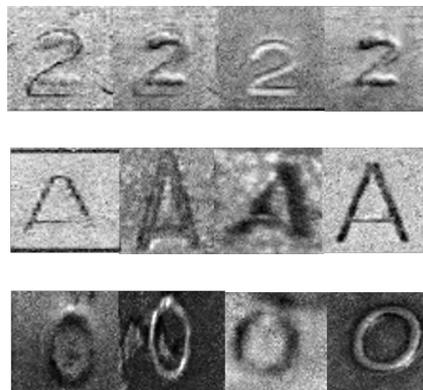
values can be attributed to the inherent complexity and diversity of the imprinted-character and -digit datasets. Certain digits/characters inherently possess more distinctive features and complex shapes, making their accurate generation challenging. Consequently, the generated samples for these digits/characters may exhibit higher FID values, indicating a greater dissimilarity with respect to the real data distribution. Conversely, numbers/characters with simpler shapes and fewer intricate details may yield lower FID values, indicating better alignment with the real data distribution. Despite the visual similarity between the generated and original images, the high FID score can be attributed to factors such as sensitivity to intra-class mode dropping, smaller sample size, and dataset characteristics [44].



**Figure 4.** Input samples: 'A', '2', and '0', selected from the original dataset shown in Figure 3.



**Figure 5.** BGAN output samples.



**Figure 6.** GAN output samples.

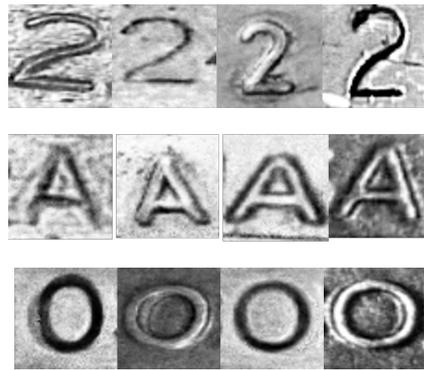


Figure 7. WGAN-GP output samples.

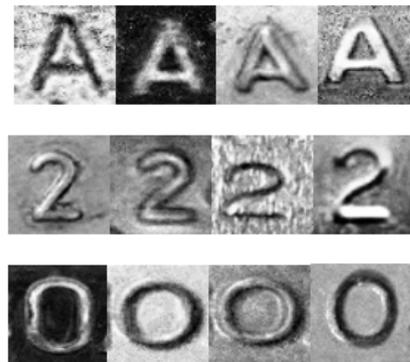


Figure 8. WGANDIV output samples.

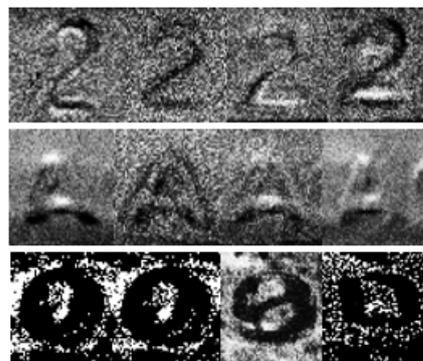


Figure 9. InfoGAN output samples.



Figure 10. LSGAN output samples.

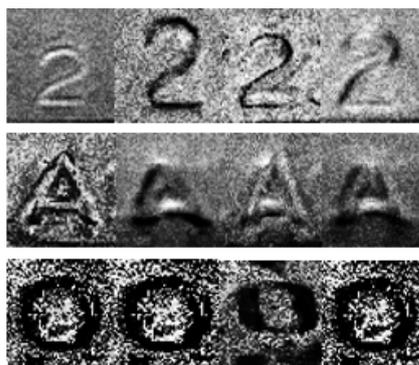


Figure 11. NSGAN output samples.

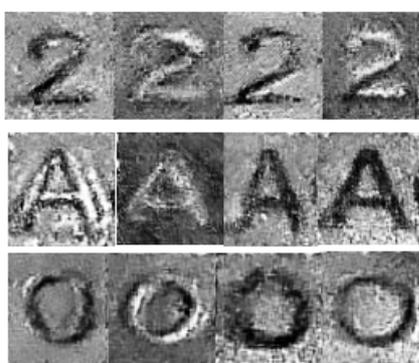


Figure 12. REL-GAN output samples.

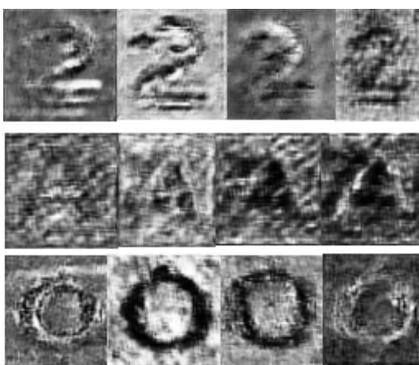


Figure 13. EBGAN output samples.

According to [44], the performance of each model is considerably influenced by the dataset, and there is no model that strictly outperforms the others. Figure 14 shows the original images from which Figures 15 and 16 are generated from. We observed that compared with other letters, certain letters, such as E, L, and I, did not exhibit considerable variations in the generated images across the top-performing GANs. This can be attributed to their simple shapes and low representation of variations in the dataset. The letters E, L, and I possess straightforward and uncomplicated structures, while other letters may have more complex curves and details. The GAN models can easily and accurately generate simpler shapes, resulting in less variation in the generated images for these letters. This finding suggests that the style or structure of each character can influence the diversity of the generated images. Figure 17 shows the selected data samples of the other letters and digits.



Figure 14. Input samples: 'I', 'L', and 'E', selected from the original dataset shown in Figure 3.

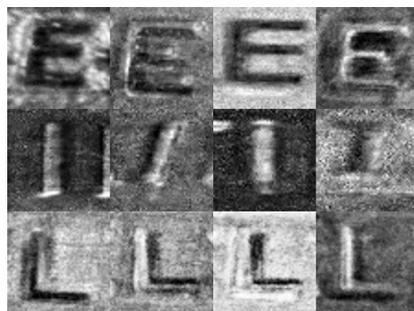


Figure 15. Images of letters I, E, and L from WGANDIV model.

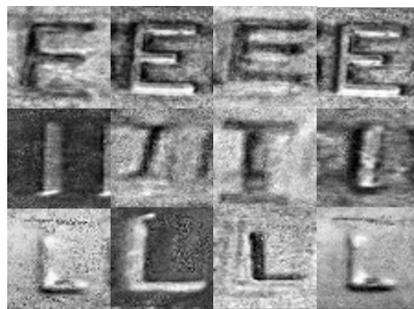


Figure 16. Images of letters I, E, and L from WGAN-GP model.

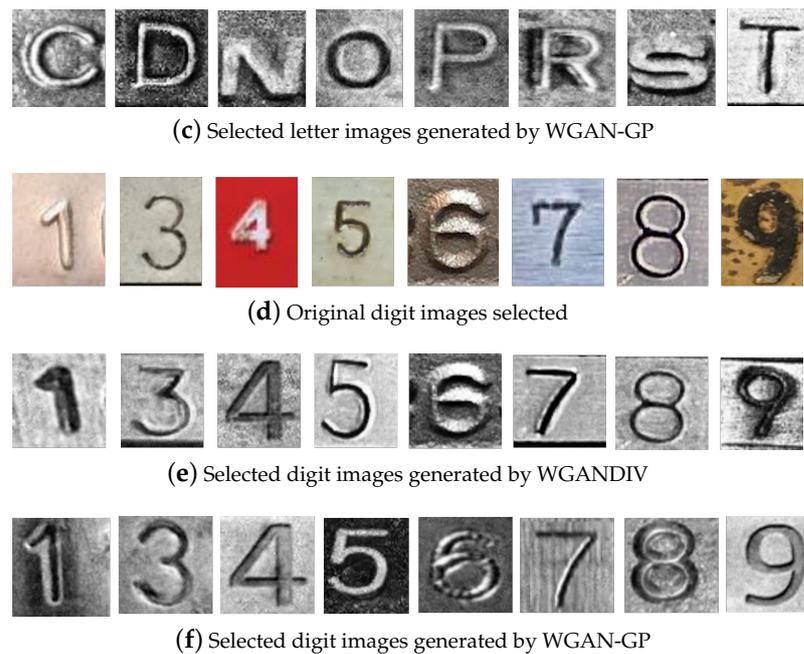


(a) Original letter images selected



(b) Selected letter images generated by WGANDIV

Figure 17. Cont.



**Figure 17.** Selected data samples of the other letters and digits.

Monitoring and analyzing the sampled loss graphs in Figures 18a–d and 19a–d (WGANDIV for ‘0’, WGANDIV for ‘2’, WGANDIV for ‘D’, WGANDIV for ‘A’; WGAN-GP for ‘1’, WGAN-GP for ‘5’, WGAN-GP for ‘D’, and WGAN-GP for ‘R’) during training provided valuable insights into the learning process of the GAN models. They revealed how the model adapted and improved over time. By analyzing these plots, we can gain a deeper understanding of the training dynamics and convergence of GAN models. The loss plots of the discriminator and generator depicting changes in the respective losses with the iterative training of the models provide valuable insights into the optimization process of GAN models. At the beginning of training, the discriminator loss is typically high due to the random initialization of the discriminator network. As the generator produces initial samples that lack resemblance to real samples, the discriminator easily distinguishes them as fake, resulting in a high discriminator loss. Simultaneously, the generator loss is also high because the generated samples fail to effectively deceive the discriminator. As training progresses, the discriminator gradually improves its discriminatory capabilities and becomes more proficient at accurately classifying real and generated samples, resulting in a decrease in the discriminator loss. The learning of the discriminator can be observed as a decrease in the slope of the loss curve, indicating the increased ability of the model to differentiate between real and generated samples. Conversely, the generator loss initially decreases as the generator learns to produce more plausible samples that can better deceive the discriminator. With backpropagation, the generator refines its parameters and adjusts its output to generate samples that progressively resemble real samples. Consequently, it becomes increasingly challenging for the discriminator to distinguish between real and generated samples, resulting in a decrease in the generator loss. The convergence of the losses indicates the optimization progress of the GAN models. Ideally, successful training can yield low values for discriminator and generator losses, indicating that the discriminator accurately classifies samples and the generator produces samples that closely resemble real ones. The convergence of the loss curves indicates that the models have reached a stable equilibrium, where the generator effectively captures the underlying data distribution.

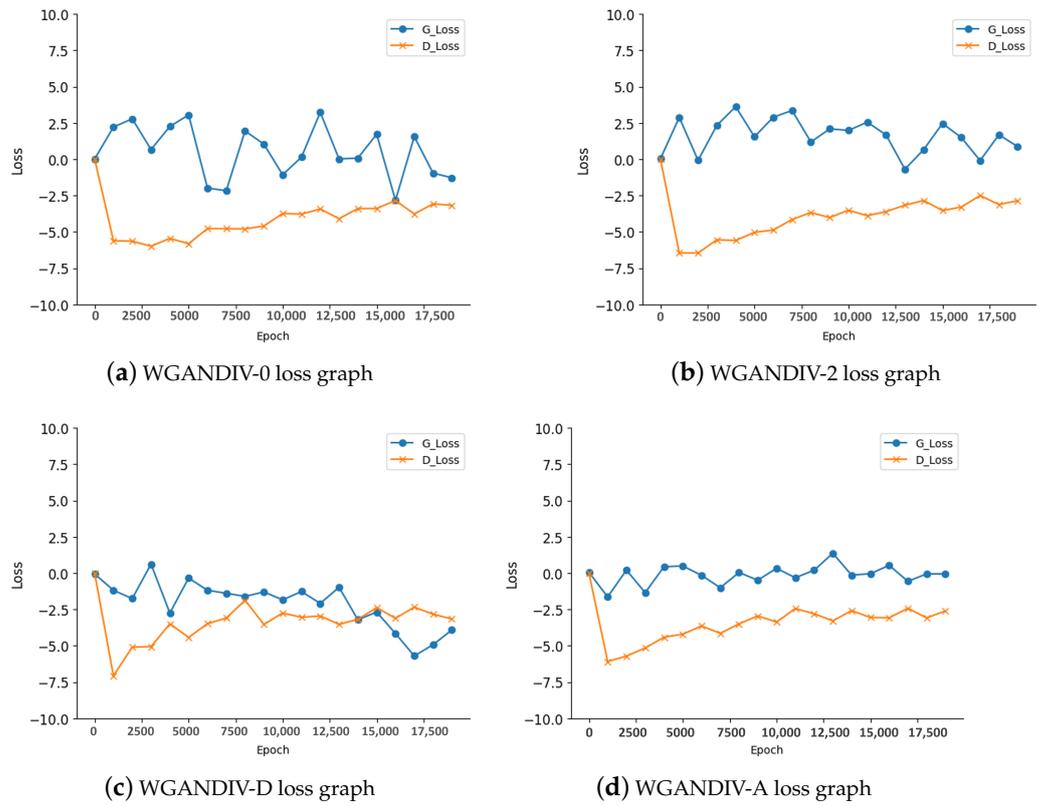


Figure 18. Loss graphs of WGANDIV.

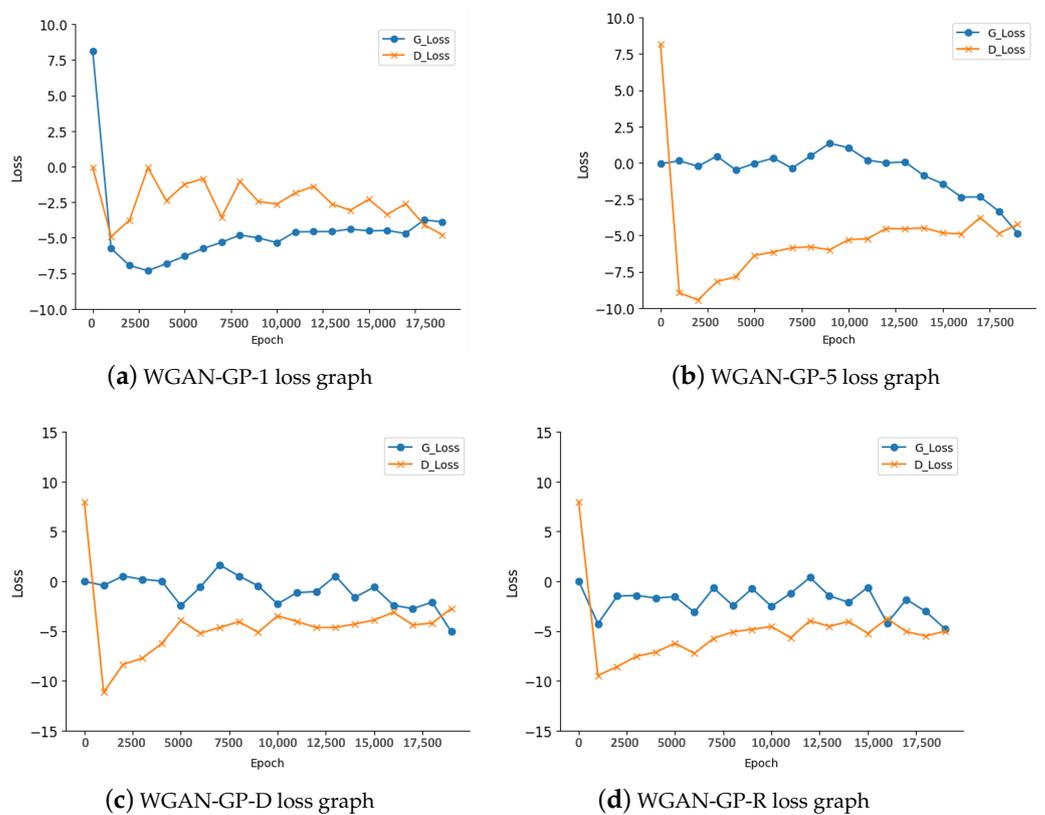


Figure 19. Loss graphs of WGAN-GP.

Based on our analysis, we determined that the optimal output for the evaluated WGAN-GP and WGAN-DIV models can be achieved around 12,500 epochs. At this point, we observed the convergence of the discriminator and generator. Notably, the GANs already exhibited promising results at approximately 12,500 epochs, indicating considerable improvements in image quality within a relatively shorter training duration. As part of our experimental evaluation, we analyzed the training duration for each of the assessed GAN models on our Ubuntu server equipped with two NVIDIA TITAN RTX GPUs. The average training time, measured in minutes, for 50,000 epochs varied across the models: WGAN (26.62), WGAN-GP (33.18), WGAN-DIV (37.84), GAN (27.42), MMGAN (24.43), NSGAN (28.21), BGAN (66.15), BEGAN (39.21), EBGAN (94.176), DRAGAN (100.56), SGAN (108.18), LSGAN (104.73), INFOGAN (96.24), and REL-GAN (107.59). These differences in training time can be attributed to several factors. Model architecture complexity, the number of parameters, and the convergence behavior are key influencers. Models with more intricate architectures and larger parameter spaces often require longer training times to achieve convergence. Additionally, GANs that introduce unique regularization techniques or novel loss functions might also require more iterations to reach an optimal balance between the generator and discriminator networks. Moreover, variations in training time can also be influenced by the computational resources available, such as the processing power of the machine used for training. These varying training durations, influenced by architectural complexity, convergence characteristics, and available computational resources, provide insights into the different demands of the evaluated GAN models.

To assess the training stability and quality of our GAN models for generating imprinted digits, monitoring the loss curves played a crucial role. With our evaluation, we closely examined the loss plots of the discriminator and generator to gain insights into the learning process and ensured effective convergence of the models. Initially, we observed high discriminator and generator losses because the models were randomly initialized and the generated samples did not closely resemble real imprinted digits. However, as training progressed, we observed a gradual decrease in the discriminator loss. This revealed that the discriminator improved its ability to accurately classify real imprinted digits from the generated ones. Simultaneously, the generator loss exhibited a downward trend, indicating that the generator was learning to produce imprinted-digit samples that closely resembled the real digits. This improvement was evident as the discriminator found it increasingly challenging to distinguish between the real imprinted digits and the generated ones. The convergence of the loss curves served as a crucial indicator of the optimization progress of our imprinted-digit GAN models. As the losses approached lower values and the curves exhibited stability, we inferred that the models were reaching a stable equilibrium. This suggested that the generator successfully captured the intricate details and style of the imprinted digits, while the discriminator became highly accurate in differentiating real imprinted digits from the generated ones. By monitoring the loss curves, we identified the potential issues during training, such as fluctuations, sudden spikes, or plateaus. These observations enabled us to address problems such as mode collapse, instability, or inadequate training. Careful analysis of the loss plots guided our decisions regarding hyperparameter tuning, regularization techniques, and architectural modifications. This iterative process helped in improving convergence and generating high-quality imprinted-digit samples that faithfully replicated the intricate details of the original engravings.

## 5. Limitations and Future Work

A notable limitation in our study is the technical challenge associated with mode collapse, a phenomenon wherein a generative model, such as a GAN, fails to capture the full diversity of the real data distribution, leading to reduced variety in the generated samples. Within the context of our research, the presence of a relatively small dataset encompassing only a few alphabets introduces the potential risk of exacerbating mode collapse. This concern informed our strategic decision to concentrate on a specific subset of characters (13 of 26 alphabet characters). By doing so, we aimed to ensure both diversity

and realism in the augmentation process, thereby mitigating the likelihood of mode collapse and its consequential negative impact on the quality and generalizability of the generated samples. Despite this limitation, we hold a strong belief in the broader applicability of our results and conclusions. Our evaluation methodology encompassed a thorough assessment of visual quality, quantitative metrics, and the practical implications of our findings. These evaluations consistently underscored the efficacy of our chosen approach in enhancing the recognition of engraved characters. Consequently, the techniques applied to the selected subset of characters offer promising avenues for broader application, substantiating the generalizability of our approach beyond the specific character set considered in this study. Future work could explore the incorporation of domain-specific knowledge into GAN models, which could significantly improve the applicability of generated images. The applicability of generated synthetic images goes beyond the maritime field. The ability to incorporate environmental and contextual factors into GAN models could potentially be applied to other industries that rely on image recognition, such as outdoor robotics, agricultural monitoring, and infrastructure maintenance.

## 6. Conclusions

In conclusion, our research has demonstrated the efficacy of GAN models in augmenting limited datasets of imprinted digits and characters for ship-character recognition. The WGAN-GP and WGAN-DIV models were able to generate diverse yet realistic digit images that are seamlessly aligned with ship-related engravings. The significance of these findings lies in their potential to significantly enhance maritime safety, operational efficiency, and security by bolstering character recognition capabilities. Our study has made significant progress in addressing data scarcity challenges in ship-character recognition. However, there are still many unexplored possibilities. Future work could explore the incorporation of domain-specific knowledge into GAN models, which could significantly improve the applicability of generated images. The applicability of generated synthetic images goes beyond the maritime field.

**Author Contributions:** A.A., J.T.S. and I.Y.J. conceived and designed the experiments; A.A. and J.T.S. performed the experiments; A.A., J.T.S., and I.Y.J. analyzed the data; A.A. and J.T.S. wrote the paper. I.Y.J. re-organized and corrected the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (No. 2021R1F1A1064345) and by the BK21 FOUR project funded by the Ministry of Education, Korea (No. 4199990113966).

**Data Availability Statement:** <https://github.com/k-sumtin/ShipMarkingsCharacterDataset.git>, accessed on 23 July 2023

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

## References

1. Western Central Atlantic Fishery Commission. *The Marking and Identification of Fishing Vessels*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017.
2. Joseph, A.; Dalaklis, D. The international convention for the safety of life at sea: highlighting interrelations of measures towards effective risk mitigation. *J. Int. Marit. Saf. Environ. Aff. Shipp.* **2021**, *5*, 1–11. [[CrossRef](#)]
3. IMO. *International Convention for the Safety of Life at Sea: Consolidated Text of the 1974 SOLAS Convention, the 1978 SOLAS Protocol, the 1981 and 1983 SOLAS Amendments*; IMO: London, UK, 1986.
4. Wawrzyniak, N.; Hyla, T.; Bodus-Olkowska, I. Vessel identification based on automatic hull inscriptions recognition. *PLoS ONE* **2022**, *17*, e0270575. [[CrossRef](#)] [[PubMed](#)]
5. Wei, K.; Li, T.; Huang, F.; Chen, J.; He, Z. Cancer classification with data augmentation based on generative adversarial networks. *Front. Comput. Sci.* **2022**, *16*, 1–11. [[CrossRef](#)]
6. Kiyoyuki dos Santos Tanaka, F.H.; Aranha, C. Data Augmentation Using GANs. *arXiv* **2019**, arXiv:1904.09135.
7. Wickramaratne, S.D.; Mahmud, M.S. Conditional-GAN based data augmentation for deep learning task classifier improvement using fNIRS data. *Front. Big Data* **2021**, *4*, 659146. [[CrossRef](#)]

8. Moon, S.; Lee, J.; Lee, J.; Oh, A.R.; Nam, D.; Yoo, W. A Study on the Improvement of Fine-grained Ship Classification through Data Augmentation Using Generative Adversarial Networks. In Proceedings of the 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju-do, Republic of Korea, 20–22 October 2021; pp. 1230–1232. [[CrossRef](#)]
9. Shin, H.C.; Lee, K.I.; Lee, C.E. Data Augmentation Method of Object Detection for Deep Learning in Maritime Image. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 463–466. [[CrossRef](#)]
10. Suo, Z.; Zhao, Y.; Chen, S.; Hu, Y. BoxPaste: An Effective Data Augmentation Method for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 5761. [[CrossRef](#)]
11. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
12. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 9759–9768.
13. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 19–21 May 2017; pp. 1–4. [[CrossRef](#)]
14. Jiao, J.; Zhang, Y.; Sun, H.; Yang, X.; Gao, X.; Hong, W.; Fu, K.; Sun, X. A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection. *IEEE Access* **2018**, *6*, 20881–20892. [[CrossRef](#)]
15. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 214–223.
16. You, A.; Kim, J.K.; Ryu, I.H.; Yoo, T.K. Application of generative adversarial networks (GAN) for ophthalmology image domains: A survey. *Eye Vis.* **2022**, *9*, 1–19. [[CrossRef](#)]
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
18. Escorcia-Gutierrez, J.; Gamarra, M.; Beleño, K.; Soto, C.; Mansour, R.F. Intelligent deep learning-enabled autonomous small ship detection and classification model. *Comput. Electr. Eng.* **2022**, *100*, 107871. [[CrossRef](#)]
19. Xu, Y.; Wang, X.; Wang, K.; Shi, J.; Sun, W. Underwater sonar image classification using generative adversarial network and convolutional neural network. *IET Image Process.* **2020**, *14*, 2819–2825. [[CrossRef](#)]
20. Starynska, A.; Easton, R.L., Jr.; Messinger, D. Methods of data augmentation for palimpsest character recognition with deep neural network. In Proceedings of the 4th International Workshop on Historical Document Imaging and Processing, Kyoto, Japan, 10–11 November 2017; pp. 54–58.
21. Wang, J.; Yu, L.; Zhang, W.; Gong, Y.; Xu, Y.; Wang, B.; Zhang, P.; Zhang, D. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 515–524.
22. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
23. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier gans. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 2642–2651.
24. Devon Hjelm, R.; Jacob, A.P.; Che, T.; Trischler, A.; Cho, K.; Bengio, Y. Boundary-Seeking Generative Adversarial Networks. *arXiv* **2017**, arXiv:1702.08431.
25. Berthelot, D.; Schumm, T.; Metz, L. Began: Boundary equilibrium generative adversarial networks. *arXiv* **2017**, arXiv:1703.10717.
26. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
27. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein GANs. *arXiv* **2017**, arXiv:1704.00028.
28. Wu, J.; Huang, Z.; Thoma, J.; Acharya, D.; Van Gool, L. Wasserstein divergence for gans. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 653–668.
29. Kodali, N.; Abernethy, J.; Hays, J.; Kira, Z. On convergence and stability of gans. *arXiv* **2017**, arXiv:1705.07215.
30. Zhao, J.; Mathieu, M.; LeCun, Y. Energy-based generative adversarial network. *arXiv* **2016**, arXiv:1609.03126.
31. Mroueh, Y.; Sercu, T. Fisher GAN. *arXiv* **2017**, arXiv:1705.09675.
32. Chen, X.; Duan, Y.; Houthoofd, R.; Schulman, J.; Sutskever, I.; Abbeel, P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. *arXiv* **2016**, arXiv:1606.03657.
33. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2794–2802.
34. Goodfellow, I. Nips 2016 tutorial: Generative adversarial networks. *arXiv* **2016**, arXiv:1701.00160.
35. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.
36. Chavdarova, T.; Fleuret, F. Sgan: An alternative training of generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 9407–9415.
37. Borji, A. Pros and Cons of GAN Evaluation Measures. *CoRR* **2018**, *179*, 41–65. [[CrossRef](#)]

38. Denton, E.L.; Chintala, S.; Fergus, R. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. *arXiv* **2015**, arXiv:1506.05751.
39. Shmelkov, K.; Schmid, C.; Alahari, K. How good is my GAN? In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 213–229.
40. Gerhard, H.E.; Wichmann, F.A.; Bethge, M. How sensitive is the human visual system to the local statistics of natural images? *PLoS Comput. Biol.* **2013**, *9*, e1002873. [[CrossRef](#)] [[PubMed](#)]
41. Zhu, X.; Vondrick, C.; Fowlkes, C.C.; Ramanan, D. Do we need more training data? *Int. J. Comput. Vis.* **2016**, *119*, 76–92. [[CrossRef](#)]
42. Sajjadi, M.S.; Bachem, O.; Lucic, M.; Bousquet, O.; Gelly, S. Assessing generative models via precision and recall. *arXiv* **2018**, arXiv:1806.00035.
43. Barratt, S.; Sharma, R. A note on the inception score. *arXiv* **2018**, arXiv:1801.01973.
44. Lucic, M.; Kurach, K.; Michalski, M.; Gelly, S.; Bousquet, O. Are GANs created equal? A large-scale study. *arXiv* **2018**, arXiv:1711.10337.
45. 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. *arXiv* **2017**, arXiv:1706.08500.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.