


Rare Data Image Classification System Using Few-Shot Learning

Juhyeok Lee and Mihui Kim * 

School of Computer Engineering & Applied Mathematics, Computer System Institute, Hankyong National University, Jungang-ro, Anseong-si 17579, Gyeonggi-do, Republic of Korea; xpdllwm95@hknu.ac.kr

* Correspondence: mhkim@hknu.ac.kr

Abstract: Advances in deep learning can address a variety of computer vision problems. In particular, deep learning has shown high performance in image processing. However, large datasets are required to train deep learning models. Previous studies have addressed the problem of data scarcity via the few-shot learning technique. However, a drawback of these studies is that large datasets are required when new tasks are performed. Hence, this study uses data augmentation techniques to address this shortcoming. Furthermore, we propose an image classification system with a few-shot learning technique that achieves high accuracy, even on rare datasets. Compared with traditional image classification models, the proposed system improves classification accuracy by approximately 18% using 100 data points.

Keywords: image classification; few-shot learning; deep learning

1. Introduction

Advances in deep learning have enabled the solution of various computer vision problems. In particular, deep learning has demonstrated high performance in image processing. However, deep learning is limited by its reliance on large, labeled image datasets for training [1].

This presents a challenge because data collection and labeling can be costly and time-consuming. Data scarcity is a problem in applications involving medical imaging, new product imaging, and art from specific artists. Various data augmentation techniques have been proposed to address this issue [2].

Data augmentation increases the amount of available data by augmenting existing data. This technology can improve the performance of deep learning image classification models [3]. However, if users apply such technology indiscriminately, the generated images may fail to accurately reflect the characteristics of the original images.

The inability to properly train the characteristics of a particular image class degrades the classification performance of a model. For rare data, the amount of actual data is limited. Hence, if data augmentation techniques are used based on actual data, the images generated via data augmentation techniques may differ significantly from natural images. Therefore, it is necessary to discriminate between images generated using data augmentation techniques and natural images [4].

This study aims to improve the classification accuracy of rare data and the classification accuracy of an image classification model by using the few-shot learning (FSL) technique. The FSL can effectively generalize a model by using a small amount of labeled data. FSL calculates a similarity score using the training data to obtain a final classification prediction result. Here, we aim to address the challenge of distinguishing between fake and real images by applying the FSL process to a real rare dataset. To verify the proposed system, we compare the classification accuracy for images generated by the augmented system with those of existing deep learning methods. In addition, we compare the classification accuracies of models with and without FSL to demonstrate the effectiveness and performance of the proposed system.



Citation: Lee, J.; Kim, M. Rare Data Image Classification System Using Few-Shot Learning. *Electronics* **2024**, *13*, 3923. <https://doi.org/10.3390/electronics13193923>

Academic Editors: Ping-Feng Pai and Krzysztof Szczypiorski

Received: 15 September 2024

Revised: 30 September 2024

Accepted: 1 October 2024

Published: 4 October 2024



Copyright: © 2024 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/>).

The remainder of this paper is organized as follows: Section 2 describes the technologies used in the proposed system. Section 3 describes a rare data image classification system that uses FSL. Section 4 provides the experimental results and an analysis of the proposed system. Section 5 concludes the paper.

2. Related Research Literature

2.1. Data Augmentation

Data augmentation improves the generalization performance of machine and deep learning models by increasing the size of the dataset via the modification or generation of additional image data. Techniques such as rotation, scaling, and symmetry are used to transform the image data. In addition to these transformation techniques, other methods, such as random cropping, flipping, and color adjustment, exist.

Data augmentation is effective when data are scarce. The collection of rare data incurs high costs and consumes a substantial amount of time. Therefore, researchers use data augmentation to increase the dataset size. Recent studies have employed generative deep learning techniques, such as generative adversarial networks (GANs) [5], to augment datasets by generating new images that are similar to existing images. This study uses data augmentation techniques to create a test dataset of generated and original images to verify the discrimination of rare data.

2.2. Convolutional Neural Networks

Convolutional neural networks (CNNs) are deep learning models [6] used in image processing and computer vision. CNNs primarily use convolutional layers to scan and extract features from images. They use small filters to scan an image and combine low-level features through convolution to understand high-level features in the image. In addition to convolutional layers, CNNs use activation functions, pooling layers, and fully connected layers to extract image features. This study used a CNN as the foundation for FSL to extract features from images.

2.3. Few-Shot Learning

FSL is a machine learning technique that can effectively learn from a few labeled data samples [7]. FSL was inspired by human learning abilities and employs meta-learning, which is a learning method that distinguishes between the known and unknown. FSL uses this meta-learning approach to make predictions using learning and query data [8]. In this study, FSL was used to classify fake and existing data using a small amount of data.

2.4. Related Works

Recently, researchers have proposed image and video classification models that use FSL to solve rare data problems.

Zhmoginonet et al. [9] proposed an FSL-based image classification model based on a transformer with a HyperTransformer. This model generates weights through high-dimensional embeddings, generates low-dimensional embeddings, delivers them to a trainable model, and learns weights to classify images. Owing to its high-dimensional embedding capabilities, the HyperTransformer model learns with little data when new tasks arise. However, if the model does not have sufficient data to learn high-dimensional embeddings, its performance degrades.

Alayrac et al. [10] proposed Flamingo, a classification model that is pre-trained using video–text pairs. The Flamingo model improved the performance of multi-model learning by analyzing the relationship between video data and text. In the learning process, FSL is used to design various models that can learn from examples and adapt to new tasks. This approach eliminates hyperparameter adjustment and demonstrates adequate performance, even under limited data conditions. Flamingo models require large-scale datasets in the pre-learning process. Therefore, when data on new tasks are insufficient, performance degrades.

Soudy et al. [11] designed an image classification model that uses FSL to solve image scene classification problems. Object detection and classification analyze image information. This generic Conv model requires a large dataset to increase performance over a specific dataset.

Previous studies [9–11] proposed a model for classifying images using machine learning and deep learning. These studies improved the performance of an image classification model by creating a pre-learning model using large datasets. However, the classification accuracy performance deteriorates when a novel task is required. Hence, this study secures several rare data by using a data augmentation model to cope with novel tasks. In addition, the use of FSL can enhance the performance of the image classification model, particularly in the context of rare data classification.

3. Proposed System

Figure 1 depicts the rare data image classification system proposed in this study, which utilizes FSL. The proposed system comprises data augmentation and image classification models. The data augmentation model is used to preprocess a dataset that is then used by the image classification model. The image classification model learns from the processed dataset and predicts the results of fake and real images using features learned through CNNs and FSL.

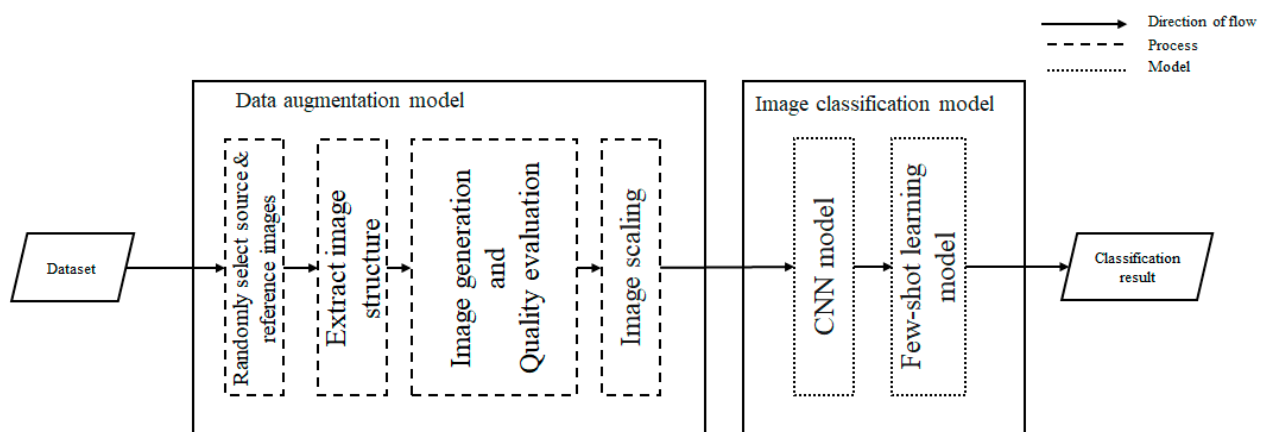


Figure 1. Proposed system flowchart.

3.1. Data Augmentation Model

Figure 2 shows a flowchart of the data augmentation model. When inputting the dataset, the data augmentation model selects the source image. Selected source images use data from other images as reference images.

Object detection is performed on the selected images to obtain the location information of objects within photos. By understanding the structures of the objects in the images, the model combines the object location information from the source image with the object information from the reference image to generate new images, thereby creating additional data.

An image structure extraction process synthesizes the source and reference images to prevent the disruption of the structural elements within them. By understanding the image structure, the model can identify the structure and shape within the images and synthesize the characteristics of the reference image with the source image, without significantly damaging the structure of the source image.

The synthesis process utilizes an encoder–decoder. The encoder encodes the source and reference images into a latent space, and the decoder reconstructs the source image using the features obtained from the encoder. In the encoder process, the Conv2D layers and activation functions extract the features. The decoder uses transposed convolutional layers that oppose the convolutional layers in the encoder to reconstruct images based on

extracted features and perform upsampling to generate images. Activation functions and batch normalization reduce the noise that may occur during image generation and minimize training loss in both the encoder and decoder layers. The structure of the generative model repeats with an encoder and a decoder to achieve an optimal loss value.

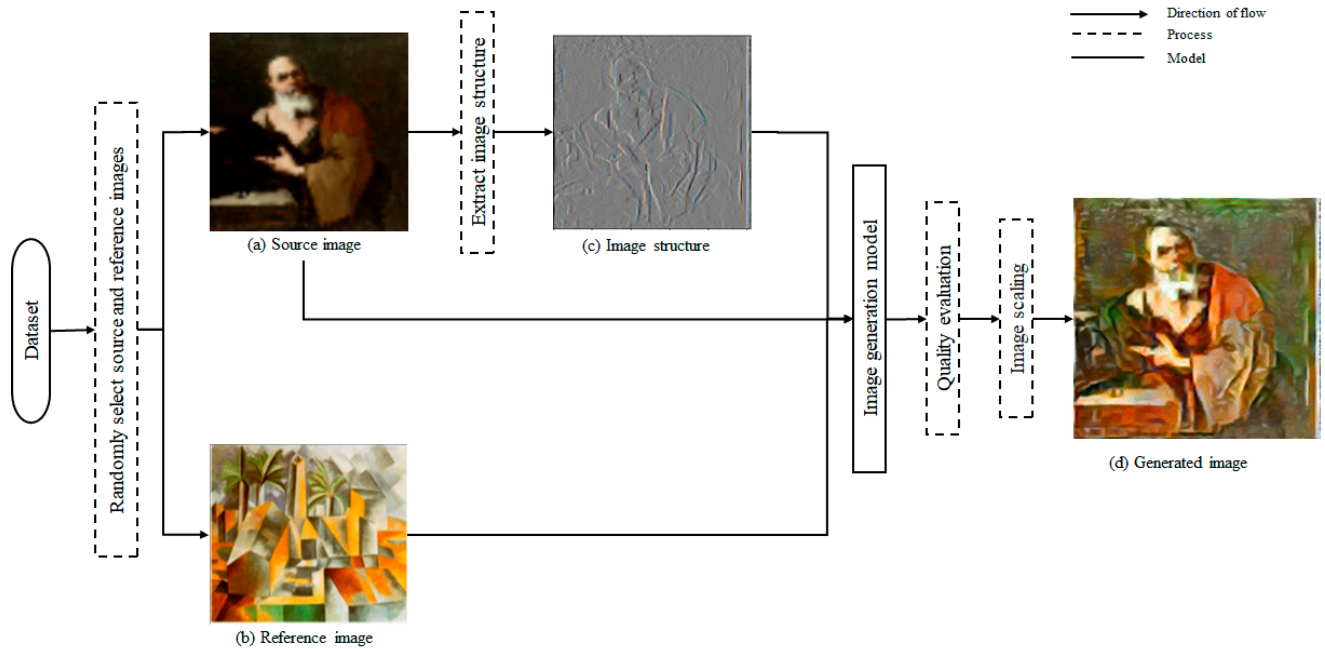


Figure 2. Data augmentation model flowchart.

The dataset obtained from the data augmentation model is then used as the test dataset for the image classification model, which is then used to distinguish between the generated and original images.

The data augmentation model obtains images from the peak signal-to-noise ratio (PSNR) [12] and Structural Similarity Index Measure (SSIM) [13] quality evaluation algorithms and preprocesses them to remove low-quality images and perform image scaling. The preprocessed data serve as the dataset for the proposed system [14]. Figure 3 shows examples of images obtained using the data augmentation model.

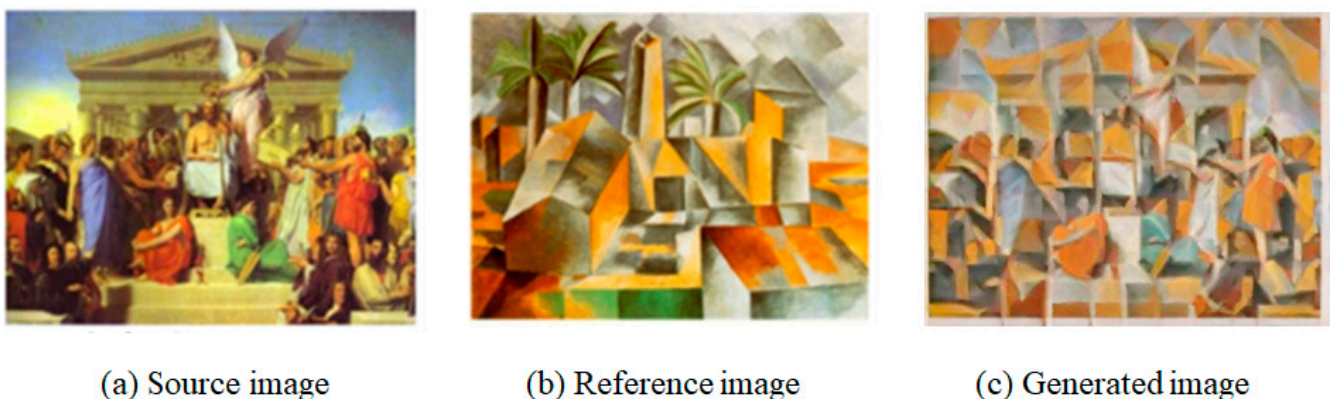


Figure 3. Example data augmentation model results.

3.2. Image Classification Model

Figure 4 shows the architecture of an image classification model. The image model consists of a CNN model and an FSL model, with the CNN model serving as the basis

for the FSL model. The CNN structure identifies features within the image by utilizing convolution, max-pooling, and flatten layers.

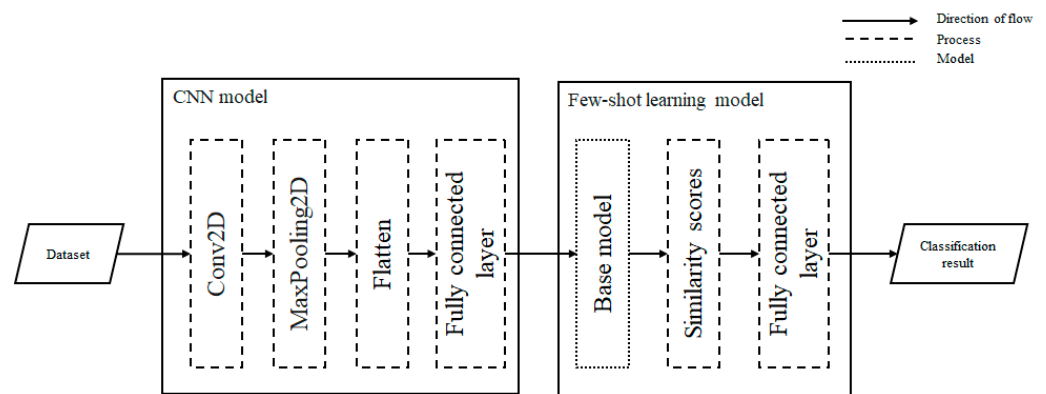


Figure 4. Image classification model flowchart.

The convolution layer captures features within an image via convolution operations for image classification. The max-pooling layer reduces the image size while accentuating the features extracted by the convolution layer. Moreover, the results from the repeated application of the convolution and max-pooling layers are converted to one-dimensional data using the flatten layer. This allows for combination with a text model, as the text data must share the same dimension [15].

The critical layers in the CNN model include Conv2D, MaxPooling2D, flatten, and dense, arranged in a repeated pattern. Using this repeated structure, the CNN model learns features and patterns within images. In addition, it performs binary classification, taking an image as the input and classifying it as natural or fake.

In the convolution layer, 32 filters of size 3×3 pixel traverse the image to extract features. The extracted features pass through the rectified linear units (ReLU) activation function, and the weights update accordingly. Then, the feature maps from the convolution layer undergo max-pooling to reduce their size. This helps prevent overfitting and improve the generalization performance of the model.

The model repeatedly applies the convolution and max-pooling layers and converts the resulting 2D feature maps into a 1D vector using the flatten layer, which is essential for input to the fully connected layer. Finally, the fully connected layer with a single neuron in the hidden layer takes the learned features as inputs to perform binary classification and obtain the final prediction.

The CNN model shown in Figure 4 serves as the base model for the FSL model. The FSL model extracts a feature vector from an input image using this base model. It calculates the means of the feature vectors and assigns them as temporary data for each class. The model then computes the cosine similarity between the feature vector and temporary data using Equation (1):

$$\text{Cosine Similarity} = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|} \quad (1)$$

Cosine similarity values range from -1 to 1 , with values closer to one indicating high similarity between the two vectors [16]. The computed similarity scores then pass through the classification layer to obtain the final prediction. The binary cross-entropy loss between the predicted and actual labels can be used to evaluate performance.

4. Experimental Results and Analysis

4.1. Experimental Data

The experimental data used painting and rare plant data from the Wikiart dataset, which contains approximately 100,000 images of artwork [17]. The Wikiart dataset includes

various types of artwork, including painting, sculpture, graphic art, and photography. The dataset was preprocessed for each artist among the artwork, and the experiment compared the accuracy of a CNN and two models using the proposed system. To confirm the relationship between the image classification performance and the amount of available data, we conducted tests with 1500 and 100 samples. Plant data were collected through web crawling, matching the number of painting data entries, including 1000, 500, and 100 samples. We then selected an appropriate number of data samples for effective image classification to evaluate performance in scenarios involving rare data [18,19]. The experimental data for image classification can be found in the GitHub repository [20].

We conducted an experiment to reduce the amount of available data and maximize the features of the rare data. We aimed to confirm the difference between the training results of the model with a large dataset of 1000 samples and the training results of the rare data. For the cases of 500 and 100 samples, we conducted experiments to compare the classification accuracy between the existing image classification model and the proposed system while gradually reducing the amount of rare data. We tested all experimental datasets with training and test data at an 8:2 ratio. In addition to the augmented system, we also tested the ArtGAN-generated [17] images provided by Wikiart. We further experimented with the ArtGAN data to verify whether the augmented model technology of the proposed system could be generalized and to evaluate the generalization performance of fake image discrimination for the generated images. Our experiment aims to determine whether the classification accuracy of the proposed system is superior to that of the existing image classification model, even in cases involving rare data. Based on the source images in Figure 5a, we used the augmented dataset obtained through the proposed system in Figure 5b and the ArtGAN dataset in Figure 5c as experimental data. The hyperparameters were chosen using Bayesian optimization [21]. We experimented with 1000 epochs, a batch size of 32, a learning rate of 0.001, and a dropout of 0.5 for both the CNN and the proposed system.

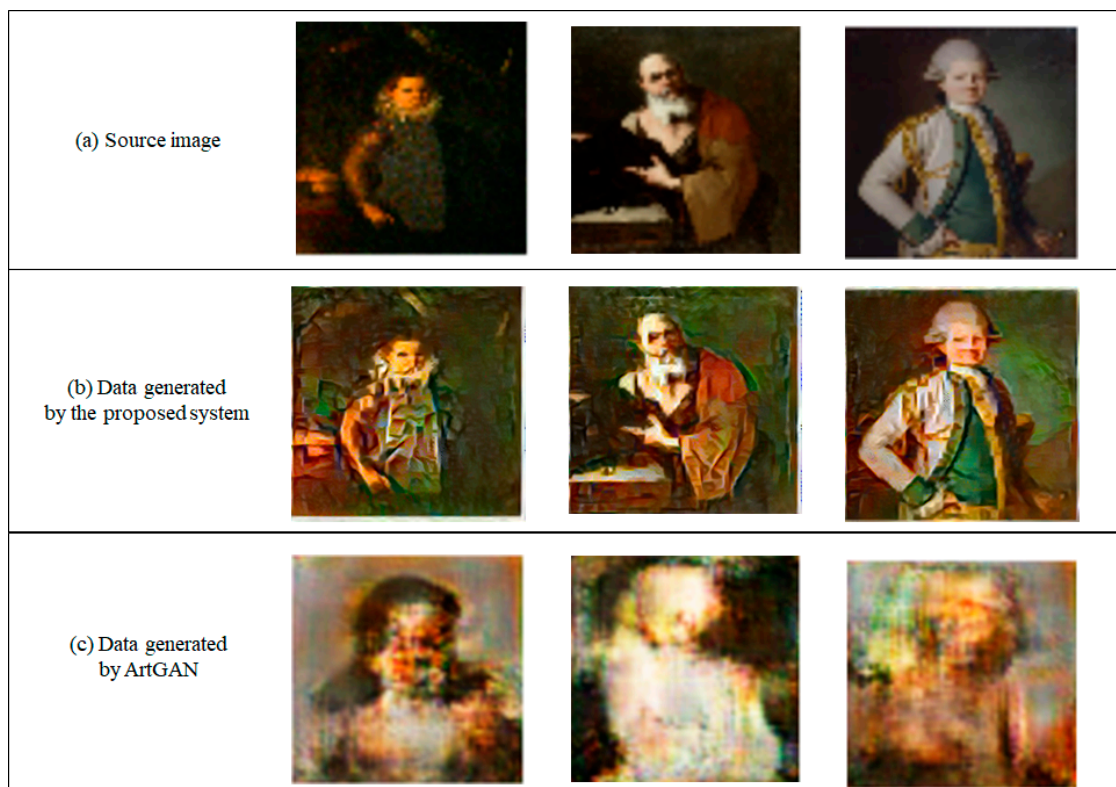
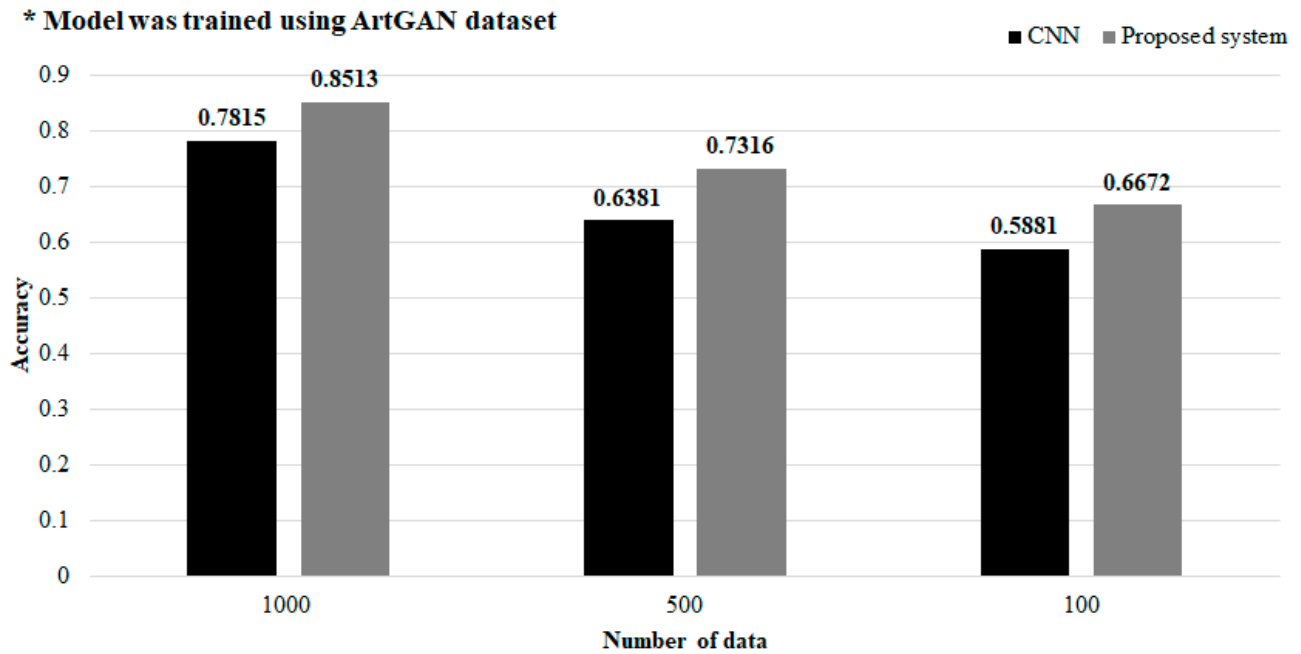


Figure 5. Example of experimental datasets.

4.2. Experimental Results

We used an image classification model and a proposed system based on a CNN to confirm the improvement in classification accuracy of the proposed system. Figures 6 and 7 show the average accuracy of the classification of fake images for painting data. Figures 8 and 9 show the average accuracy of the classification of fake images for plant data. Tables 1 and 2 show the experimental results of 100 data entries augmented with the proposed system. We used the precision, recall, and F1 scores, in addition to the accuracy of the painting and plant data, as evaluation metrics for the image classification model.



Number of data	Classification model	
	CNN	Proposed system
1000	0.7815	0.8153
500	0.6381	0.7316
100	0.5881	0.6672

Figure 6. ArtGAN painting data classification accuracy by number of data samples.

Table 1. Accuracy, recall, precision, and F1 score values of 100 painting data entries.

Method	Accuracy	Recall	Precision	F1 Score
CNN	0.5876	0.5741	0.5803	0.5978
Proposed system	0.6781	0.6812	0.6934	0.7196

Table 2. Accuracy, recall, precision, and F1 score values of 100 plant data entries.

Method	Accuracy	Recall	Precision	F1 Score
CNN	0.5881	0.5614	0.5716	0.5831
Proposed system	0.5876	0.6763	0.6901	0.7089

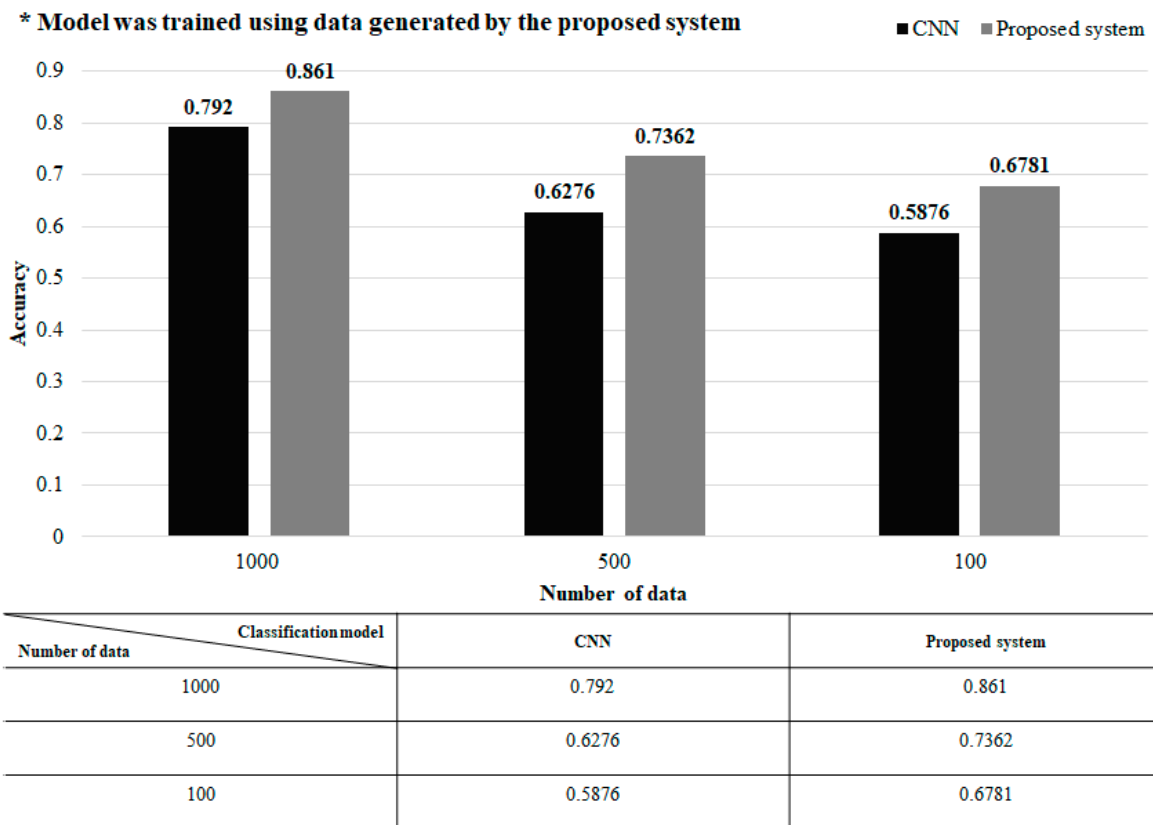


Figure 7. Painting data generated by the proposed system classification accuracy by number of data samples.

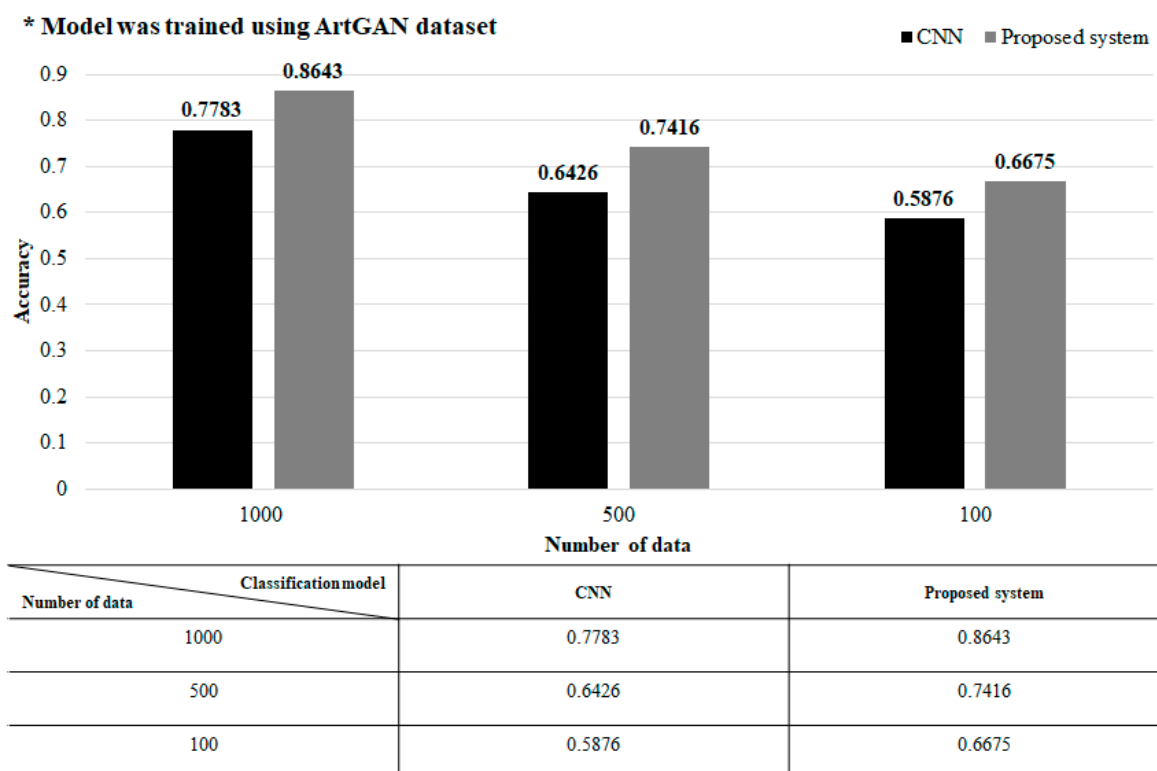


Figure 8. ArtGAN plant data classification accuracy by number of data samples.

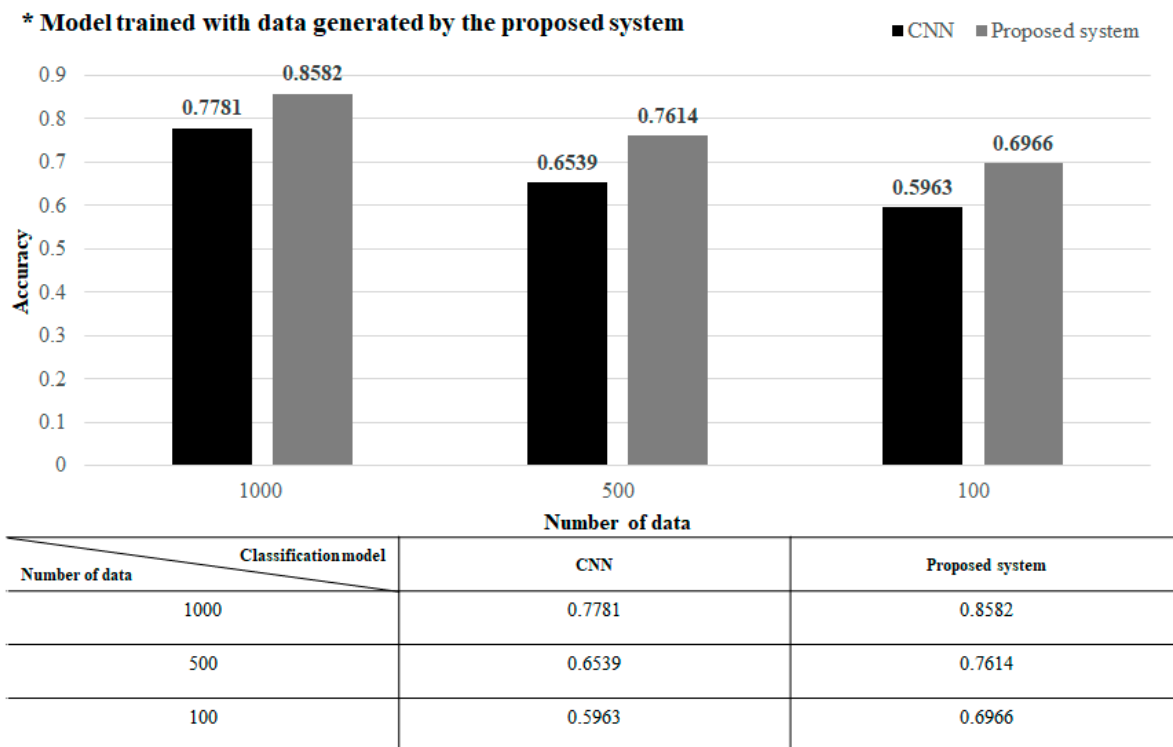


Figure 9. Plant data generated by the proposed system classification accuracy by number of data samples.

To validate the data, we verified the validity of the image classification model from the ArtGAN dataset [17], augmented data from the Wikiart dataset [17], and proposed dataset supplemented by the proposed system in Figure 5b.

As shown in Figures 6 and 7, the proposed system improved the average classification accuracy of the 1000 painting data samples during the experiment by approximately 8%, compared with the existing image classification model (i.e., the CNN). In the case of 100 samples, which had the least amount of data, the proposed system achieved an average classification accuracy of approximately 15% higher than that of the CNN.

As shown in Figures 8 and 9, in the case of the 1000 plant data samples, the proposed system was approximately 9% higher than the average classification accuracy of the CNN. In the case of 100 samples, which had the least amount of data, the proposed system achieved an average classification accuracy that was approximately 16% higher than that of the CNN.

As shown in Tables 1 and 2, the proposed system showed a reproduction rate that was approximately 19% higher than that of the CNN. The precision of the proposed system was also approximately 20% higher than that of the CNN. Through reproduction and precision, the prediction performance of the images classified by the proposed system improved to match that of the CNN. The F1 score showed an improvement of approximately 22% for the performance of the proposed system, compared with that of the CNN. Hence, based on the reproducibility, precision, and F1 score, the problem response performance improved upon that of the CNN, even in cases in which disproportionate data problems were present.

5. Conclusions

In this study, we propose a rare data image classification system using FSL. Dataset design processes for rare data incur considerable time and costs. We use data augmentation technology to solve this problem; however, it poses a challenge in distinguishing between generated fake images and existing authentic images. To overcome this limitation, we increase the classification accuracy of the fake and authentic photos using FSL techniques.

The classification accuracy of rare data is compared using the Wikiart artwork dataset, which shows the distinct characteristics of each image among many rare data samples. Experiments show that the lower the number of data points, the higher the classification accuracy of the proposed system, compared with traditional image classification models. In addition, we will study different FSL techniques related to the proposed system and system optimization. Furthermore, we will examine how the proposed system can be applied to additional types of rare data, including text and audio.

Author Contributions: J.L. and M.K. developed the proposed system. J.L. performed the experiments and evaluated the proposed system. M.K. supervised the design and development of the system proposed in this work and guided this work as the corresponding author. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no external funding.

Data Availability Statement: The experimental data for image classification is available at <https://github.com/juhyeok99/few-shot-learning-for-rare-data> (accessed on 30 September 2024).

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

References

1. Kwon, H.; Kim, Y.C. Adversarial case technology trends for deep learning models. *Inst. Informat. Secur. Cryptol.* **2021**, *31*, 5–12.
2. Tang, H.; Xu, D.; Sebe, N.; Wang, Y. A survey on multimodal deep learning for image synthesis: Applications, methods, datasets, evaluation metrics, and results comparison. *IEEE Access* **2020**, *8*, 108–120.
3. Ayub, A.; Kim, H. GAN-Based Data Augmentation with Vehicle Color Changes to Train a Vehicle Detection CNN. *Electronics* **2024**, *13*, 1231. [[CrossRef](#)]
4. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
5. Popuri, A.; Miller, J. Generative Adversarial Networks in Image Generation and Recognition. In Proceedings of the International Conference on Computational Science and Computational Intelligence, Las Vegas, NV, USA, 13–15 December 2023; pp. 1294–1297. [[CrossRef](#)]
6. Li, R.; Zhang, W.; Suk, I.; Wang, L.; Li, J.; Shen, D.; Ji, S. Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014*; Springer: Boston, MA, USA, 2014; pp. 305–312. [[CrossRef](#)]
7. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 4080–4090.
8. Wang, J.-H.; Le, P.T.; Jhou, F.-C.; Su, M.-H.; Li, K.-C.; Chen, S.-L.; Pham, T.; He, J.-L.; Wang, C.-Y.; Wang, J.-C.; et al. Few-Shot Image Segmentation Using Generating Mask with Meta-Learning Classifier Weight Transformer Network. *Electronics* **2024**, *13*, 2634. [[CrossRef](#)]
9. Zhmoginov, A.; Sandler, M.; Vladymyrov, M. HyperTransformer: Model generation for supervised and semi-supervised few-shot learning. In Proceedings of the International Conference on Machine Learning, Baltimore Convention Center, Baltimore, MD, USA, 17–23 July 2022.
10. Alayrac, J.-B.; Donahue, J.; Luc, P.; Miech, A.; Barr, I.; Hasson, Y.; Nguyen, P.; Ring, R.; Orhan, A.; Raphael, P.; et al. Flamingo: A visual language model for few-shot learning. In Proceedings of the Advances in Neural Information Processing Systems 35, New Orleans Convention Center, New Orleans, LA, USA, 28 November–9 December 2022.
11. Soudy, M.; Afify, Y.M.; Badr, N. GenericConv: A Generic Model for Image Scene Classification Using Few-Shot Learning. *Information* **2022**, *13*, 315. [[CrossRef](#)]
12. Wang, Y.; Li, J.; Lu, Y.; Fu, Y.; Jiang, Q. Image quality evaluation based on image weighted separating block peak signal to noise ratio. In Proceedings of the IEEE International Conference on Neural Networks and Signal Processing, Nanjing, China, 14–17 December 2003; Volume 2, pp. 994–997. [[CrossRef](#)]
13. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
14. Lee, H.; Kim, H. Object edge-based image generation technique for constructing large-scale image datasets. *J. IKEEE* **2023**, *27*, 280–287.
15. Choi, H.; Choi, J.; Min, H.; Chung, H.; Ahn, J. Development of de-noised image reconstruction technique using Convolutional Autoencoder for fast monitoring of fuel assemblies. *Nucl. Eng. Technol.* **2021**, *53*, 888–893. [[CrossRef](#)]
16. Nguyen, V.; Bai, L. Cosine Similarity Metric Learning for Face Verification. In *Computer Vision—ACCV 2010*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 709–720. [[CrossRef](#)]
17. Saleh, B.; Elgammal, M. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv* **2015**, arXiv:1505.00855.

18. Hariharan, B.; Girshick, R. Low-shot visual recognition by shrinking and hallucinating features. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3018–3027.
19. Li, F.-F.; Fergus, R.; Perona, P. One-shot learning of object categories. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 594–611. [[CrossRef](#)] [[PubMed](#)]
20. GitHub Repository. Available online: <https://github.com/juhyeok99/few-shot-learning-for-rare-data> (accessed on 30 September 2024).
21. Jia, W.; Chen, X.-Y.; Zhang, H.; Xiong, L.-D.; Hang, L.; Deng, S.-H. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.

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.