Data Augmentation with Synthesized Damaged Roof Images Generated by GAN

Koki Asami
Kyoto University
asami@dimsis.dpri.kyoto-u.ac.jp

Shono Fujita
Kyoto University
fujita@dimsis.dpri.kyoto-u.ac.jp

Kei Hiroi
Kyoto University
hiroi@dimsis.dpri.kyoto-u.ac.jp

Michinori Hatayama
Kyoto University
hatayama@dimsis.dpri.kyoto-u.ac.jp

ABSTRACT

The lack of availability of large and diverse labeled datasets is one of the most critical issues in the use of machine learning in disaster prevention. Natural disasters are rare occurrences, which makes it difficult to collect sufficient disaster data for training machine learning models. The imbalance between disaster and non-disaster data affects the performance of machine learning algorithms. This study proposes a generative adversarial network (GAN)-based data augmentation, which generates realistic synthesized disaster data to expand the disaster dataset. The effect of the proposed augmentation was validated in the roof damage rate classification task, which improved the recall score by 11.4% on average for classes with small raw data and a high ratio of conventional augmentations such as rotation of image, and the overall recall score improved by 3.9%.

Keywords

disaster response, generative adversarial networks, data augmentation, damage classification.

INTRODUCTION

Deep learning, as a subset of machine learning, has attracted attention owing to its performance and variety of applications. Recently, machine learning has been applied to disaster prevention (Miyamoto et al., 2020), which prevents or mitigates damage from natural disasters, such as earthquakes and floods, and quickly recover from damage. For example, disaster occurrence (Anand and Miura, 2021; Lee et al., 2020), damage scale (Chaurasia et al., 2019; Fujita and Hatayama, 2021), and damage detection (Miyamoto and Yamamoto, 2020) to gather the situation of damage after the disaster.

The great success of deep learning is not only because of the improvement of its architecture, but also the amount and variety of available training data. A large amount of training data allows deep learning models to learn the features of the data and produce the desired results. The recent trend of digital transformation (Dx) has enabled more data to be digitalized and become available as big data (Miklosik and Evans, 2020). This trend has accelerated the use of deep-learning methods. However, data collection remains a critical bottleneck in several fields (Roh et al., 2021). In particular, the collection of a sufficient amount of natural disaster data for deep learning is difficult because of the following reasons.

- Generally, natural disasters do not occur frequently; thus, compared with non-disaster data, there are fewer opportunities to collect disaster data.
- Because natural disasters are rare, the development of information technology is relatively fast. As a result, the data format used in the previous disaster may be obsolete, making the data collection difficult.
- Even when a natural disaster occurs, it is highly likely that it will destroy essential infrastructure, which interferes with data collection in the field.
- Moreover, because safety is a top priority, the number of workers available for data collection is limited.

Commented [A1]: **Tip:** Collocations refer to combinations of words that are used together very often. These combinations are often described by native-English speakers as sounding "natural" or "right."

For example,
Original: We arrived *on* the same conclusion.
Revised: We arrived *at* the same conclusion.

Original: The differences were *cleanly* visible.
Revised: The differences were *clearly* visible.

Commented [A2]: **Tip:** When two complete sentences are combined using a comma it leads to an error called a "comma splice" – a type of run-on sentence.

This can be corrected in one of three ways:
- replacing the comma with a semicolon,
- splitting the clauses into two sentences using a period,
- adding a coordinating conjunction such as and, or, nor, for, but, etc., after the comma.

For example,
Original: I went to the market, I met John.
Revision 1: I went to the market; I met John.
Revision 2: I went to the market. I met John.
Revision 3: I went to the market, and I met John.

Commented [A3]: To avoid gender-biased language, consider using 'workers', 'work force', 'staff', 'labour/labor', or 'personnel'.

Because non-disaster data are relatively easy to collect, disaster datasets tend to contain more non-disaster data than disaster data. A data imbalance leads to other critical problems when training the models. When there is a data imbalance in the training data, models tend to overclassify the majority class because of the increased prior probability (Johnson and Khoshgoftaar, 2019). Consequently, minor data classes are misclassified more frequently than majority data classes. In addition to the neural-network-based model, data imbalance is another issue (Anand et al., 1993). Training neural network-based models with imbalanced data affects the gradient components. The gradient component of the minor class is shorter than the gradient component of the majority class. This indicates that the majority class is the principal component used for updating the weights of the model. This reduces the loss of the majority class in the early epochs but increases the loss of the minor class. This causes slow convergence in the training. Generally, deep learning models require at least 5,000 labeled data per class to achieve sufficient performance, and 10,000,000 data per class to achieve performance comparable to or better than humans (Goodfellow et al., 2016). The cost of correcting a large amount of data is extremely high. If the dataset is small, machine learning models learn the features that exist only in the training data and are not common in the real world. This phenomenon is known as overfitting. When a model is overfitted to the training data, the generalizability of the model is lost, and its performance is significantly reduced.

To address overfitting and data imbalance issues, data augmentation techniques are commonly used to increase the amount of data. Simple image transformations, such as rotation and mirroring, or image processing, such as color jittering, are known to be treated as different data samples from the original data, which allows a tenfold increase in the number of samples in a dataset (Krizhevsky et al., 2012). However, these conventional data augmentation methods produce only a limited variety of alternative data (Antoniou et al., 2018). Therefore, generative adversarial network (GAN)-based data augmentation has recently attracted considerable attention. In GAN-based data augmentation, conditional GAN (Mirza and Osindero, 2014), which adds a condition vector to the generator as an input to regulate the output, is widely used, particularly for segmentation tasks in the medical field (Bowles et al., 2018; Sandfort et al., 2019; Wickramaratne and Mahmud, 2021). Other studies (Frid-Adar et al., 2018; Maeda et al., 2021; Moradi et al., 2018) used nonconditional GANs to augment training data with significant results in both classification and object detection tasks.

Commented [A4]: Introducing this abbreviation before it is used on its own.

Commented [A5]: Tip: Conciseness refers to writing using the fewest words required to convey a message clearly. This is particularly important in academic writing because most publications set a word count limit for submissions. Further, wordiness can muddle the message being conveyed. Therefore, it's best to ensure conciseness to create more impactful writing.

To ensure conciseness, it is often preferable to use a direct verb rather than using the noun form of that verb.

For example,
Original: We *performed an evaluation* of all methods used previously.
Revision: We *evaluated* all methods used previously.

This study addresses the imbalance of collected disaster data, which is one of the most critical problems when machine learning is used for disaster prevention. The application considered in this study is the classification of an earthquake-damaged roof (Fujita and Hatayama, 2021), which aims to automate the calculation of the roof damage rate for issuing a disaster certificate.

The remainder of this paper is organized as follows. Section 2 discusses the literature related to this study. Section 3 describes the approaches used in this study. The results and discussion are presented in Section 4. Finally, Section 5 presents the conclusions of this study.

RELATED WORKS

Generative Adversarial Networks

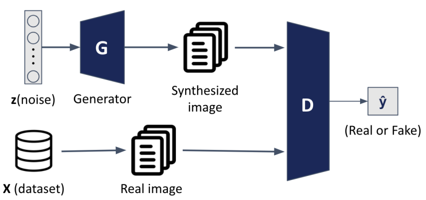

Figure 1. The structure of generative adversarial networks

Commented [A6]: Because the abbreviation for this was introduced earlier, only the abbreviation is sufficient.

GANs are a type of generative model that can generate new data based on extracted features from training data (Goodfellow et al., 2014). A GAN is composed of two neural networks: a generator (G) and a discriminator (D) (Figure 1). The generator is a generative model that captures the data distribution and generates a new data instance using the latent variable Z, which can be random noise from a uniform or normal distribution. The discriminator is a discriminative model that evaluates the authenticity of the generated data by estimating the probability that the input data comes from the training data rather than the generator's data. By training these two networks simultaneously, the generator learns to generate new data that are close to the training data distribution, which is more realistic, and the discriminator learns to improve the accuracy of distinguishing the input sample as real or fake data from the generator. This means that GAN training is a two-player minimax game in which the objective is to determine the Nash equilibrium. The cost value is calculated as follows:

$$V(D, G) = \min_{G} \max_{D} \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_{\text{generated}}(z)}[1 - \log D(G(z))]$$

where $D(x)$ represents the score when the discriminator recognizes real data and 1 - $D(G(z))$ represents the score when the discriminator recognizes fake data. The discriminator is trained to maximize the probability of correctly recognizing data as real or fake, while the generator is trained to generate realistic data to minimize $\log(1 - D(G(z))$, which is the score at which the discriminator correctly recognizes the generated data.

Because GAN can generate realistic high-quality images, it has been applied to many practical applications (Jabbar and Omar, 2020), such as simple image generation tasks (Hamada et al., 2018; Kodali et al., 2017), image synthesis tasks (Dolhansky and Ferrer, 2017; Li et al., 2018; Wu et al., 2019), and super-resolution tasks (Ledig et al., 2017; Xu et al., 2021). In addition to image generation, GAN has been used to generate text (Lin et al., 2018) and music (Mogren, 2016). Although GAN has achieved a breakthrough in generative models in terms of quality, challenges remain in GAN training (Jabbar et al., 2020). Mode collapse is the most critical problem associated with GAN training, which occurs when the generator produces outputs that are all the same or very similar. Mode collapse occurs because the discriminator evaluates generated images based on quality rather than diversity. Therefore, the generator succeeded in deceiving the discriminator with less diversity data. To address this problem, a progressively growing GAN (PG-GAN) (Karras et al., 2018) adopted a multi-scale-based GAN architecture that stabilizes training and avoids mode collapse by training both the generator and discriminator from lower resolution to higher resolution. However, this multi-scale-based GAN is independent of each other, resulting in the generation of features that should be shared within a generator, with no changes from module to module. StyleGAN2 reconstructs its architecture to avoid this problem (Karras et al., 2020b). Details of StyleGAN2 are discussed in the following section. In this study, StyleGAN2 was used to generate high-quality synthetic damaged-roof images.

Commented [A7]: **Tip:** Repeated article error refers to the unnecessary repetition of the same article in a series or list.

For example,
Original: We verified the samples using *the* source image, *the* original image, and the final image.
Revision: We verified the samples using *the* source image, original image, and final image.
It's fine to place the article before the first item in the list only.

Generative Adversarial Networks for data augmentation

In principle, deep learning models make predictions based on training data. In other words, deep learning models cannot perform well when they encounter new features. Therefore, to avoid overfitting and achieve high performance in the real world, deep learning models require a diverse and general dataset (Shorten and Khoshgoftaar, 2019). Many studies (DeVries and Taylor, 2017; Inoue, 2018; Moreno-Barea et al., 2018; Taylor and Nitschke, 2017) have been conducted on data augmentation to increase the amount of data by applying image processing, GAN-based data augmentation is one of them. GANs have gained popularity as data augmentation applications because of their impressive performance (Shorten and Khoshgoftaar, 2019). In fact, GAN-synthesized data can provide add to the information in a dataset (Bowles et al., 2018). GAN-based data augmentation is a potential solution to the lack of large and diversified datasets. GAN-based data augmentation is popular in the medical field (Bowles et al., 2018; Sandfort et al., 2019; Wickramaratne and Mahmud, 2021), which is another field that suffers from the lack of large datasets. For example, cycle GAN (Zhu et al., 2020) was used to improve the generalizability of CT image segmentation tasks (Sandfort et al., 2019). Another field that studies GAN-based augmentation is civil engineering, which is more closely related to disaster prevention. For example, Gao et al. (2019) showed that the use of DCGAN-based augmentation (Radford et al., 2016) enhances the performance of the damage state classification model.

Commented [A8]: Because this is the first time this abbreviation has been used, it must be written in full.

Commented [A9]: Note that I have edited this in-text citation.

Commented [A10]: The abbreviation DC has not been introduced.

Image-based Damage Classification in Disaster Prevention

As mentioned in the previous section, several studies in the disaster prevention field have applied machine learning. Building damage assessment automation, a popular application of machine learning in disaster prevention, is expected to alleviate manpower shortages and streamline the process of receiving support. Fujita et al. (2017) proposed using convolutional neural networks (CNNs) for binary classification of washed-away buildings from aerial images. Liu et al. (2021) proposed a more advanced system to locate individual buildings using aerial images and classify the damage into four levels: no damage, minor damage, major damage, and total damage. While Fujita et al. (2017) and Liu et al. (2021) are studies that support disaster management in general, Fujita and Hatayama (2021) proposed a system for automating disaster certificate issuance, which is a more concrete and practical objective. Fujita and Hatayama (2021)'s system classifies the roof damage rate into five levels and calculates the damage rate of the roof of a building from an aerial image for automating disaster certificate issuance. A disaster certificate is used to determine the content of support for victims.

Table 1 lists the data augmentation methods used in the literature. Fujita et al. (2017) and Fujita and Hatayama

Commented [A11]: **Consistency:** Different styles have been used when citing tables in the text.
'table(s)' [1 time]
'Table(s)' [3 times]
Please pick one style and use it consistently throughout the text.

(2021) used geometric transformations, such as vertical and horizontal flipping and rotation, as data augmentation. Geometric transformation is a widely used augmentation method owing to its low implementation cost. Additionally, it is common to use aerial images for building damage detection models that contain various angles and positions of buildings; therefore, a geometrical transformation is necessary. However, the variety of building styles cannot be augmented by geometrical transformations. According to Bowles et al. (2018), synthesized data from a GAN can provide additional information from a dataset. Tilon et al. (2020) proposed a GAN-based building damage detection model, but the main purpose of the model was not to augment data but to directly integrate it into an anomaly detection model. Existing studies on post-disaster building damage detection models only use augmentation with geometric transformation, and GAN-based augmentation has not been addressed (Fujita et al., 2017; Fujita and Hatayama, 2021; Liu et al., 2021). Therefore, this study proposes GAN-based data augmentation, which is expected to extract new features from small-sized datasets to improve the model's robustness.

Table 1. Summary of augmentation methods in the literature

	Data Augmentation	Flipping	Rotating	GAN-based
Liu et al.	Not mentioned	Nan	Nan	Nan
Fujita et al.	Yes	Yes	No	No
Tilon et al.	Not mentioned	No	No	No
Fujita and Hatayama	Yes	Yes	Yes	No

For disaster prevention, where practicality is vital, it is important to develop a system with a specific purpose. Therefore, this study addresses the study by Fujita and Hatayama (2021), which is one example of the lack of disaster data, as an application to verify the effects of the proposed GAN-based data augmentation.

METHOD

Overview

The flow of this study is illustrated in Figure 2. Note that for the fairness of comparing the classification scores, this study strictly follows the training flow proposed by Fujita and Hatayama (2021). In the first step, the StyleGAN2 model was trained using the damaged-roof image dataset (Fujita and Hatayama, 2021). Next, the generated images were divided into roof segments and annotated for use in the training classification model. Annotated data were added to the original dataset. Finally, the damage classification model was trained using a dataset with StyleGAN2-based augmentation.

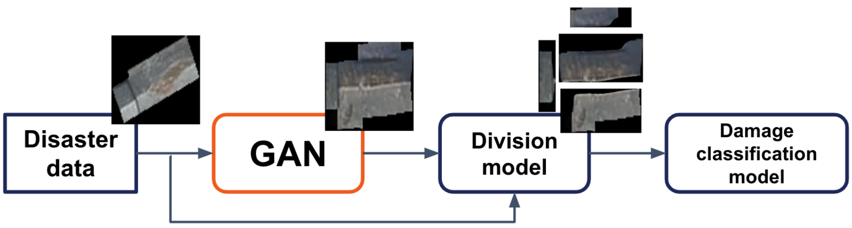

Figure 2. The flow of the study

Dataset

In this study, the damaged-roof image dataset from the 2016 Kumamoto Earthquake in Japan was used to generate synthesized images and classify the roof damage rate. This earthquake had a maximum magnitude of 7.3, making it one of the largest in Japan and causing severe damage to both humans and buildings. The original data are orthophoto map data from aerial photographs taken by the Geospatial Information Authority of Japan shortly after the earthquake occurred. Subsequently, Fujita and Hatayama (2021) developed a trimming algorithm that uses both the orthophoto map and building polygons from geospatial information data to trim individual buildings from the orthophoto map. The trimmed dataset contained 714 damaged and 530 undamaged images. The size of each

Commented [A12]: **Tip:** The expression "et alia," which means "and others" is used in written text when referring to published research with multiple authors. The first author's name is followed by the abbreviated form, "et al." which is a concise alternative for "and co-workers" or "and colleagues."

For example,
Original: Zhang and colleagues also investigated the efficacy of this treatment.
Revised: Zhang et al. also investigated the efficacy of this treatment.

Commented [A13]: Kindly double check if abnormal was the intended word here.

image was 256 × 256 pixels. In this study, the dataset was first subjected to data augmentation and image processing, and then annotated to train the damage-rate classification model, as detailed in the following sections. A summary of the datasets is presented in Table 1. Note that the number of no-damage roof segment images in the original data was reduced by Fujita and Hatayama (2021) to moderate the imbalance of data, and damaged roof segment images are augmented by mirroring or rotating at 90°, 180°, and 270°, except for the data in the -25 damage category, which has a relatively sufficient amount of data.

Table 2. Summary of dataset

	Raw Dataset	**Raw data with simple augmentation**	**Raw data with StyleGAN2-based augmentation**	**Test Data**
No Damage	2221	1103	2221	242
-25 Damage	1342	1342	1613	52
25-50 Damage	436	872	747	40
50-75 Damage	248	992	536	22
75- Damage	145	1160	563	16

Generation of synthesized images

In this study, StyleGAN2 (Karras et al., 2020b) was used to generate the synthesized damaged roof images. StyleGAN2 shares the same basic structure as StyleGAN (Karras et al., 2019). StyleGAN has been drawing considerable attention because of its ability to generate highly accurate images of people and cars that can even deceive humans. The structure of the generator is the main improvement over conventional GANs proposed by Karras et al., (2019). In contrast to a conventional GAN, the generator of a StyleGAN is composed of two networks: a mapping network and a synthesis network. The mapping network is composed of eight fully connected layers. The aim of this network is to use a nonlinear transformation to project a latent variable Z (a vector of random numbers) into a multi-dimensional feature space W of style information. This trick reduces the entanglement of the latent space; thus, the projected feature space W becomes smooth and a specific feature is mapped to a variable w, where $w \in W$, for example, w_1 represents the size of the roof and w_2 represents the number of roof segments. Similar to PG-GANs, the synthesis network is composed of several modules that generate the synthesized image from low to high resolution to stabilize the generation of high-resolution synthesized images. However, the synthesis network does not generate synthesized images from a latent variable Z or projected feature space Wbut from a constant tensor with a size of 4 × 4 × 512. Each module of the synthesis network receives two additional inputs: variable $w$and random noise. A variable w determines the global features, such as the shape of the roof, while random noise determines the local features, such as the texture of the roof. StyleGAN generates the stunning quality of the synthesized images, but there is still room for improvement. Karras et al. (2020b) analyzed the generated images from StyleGAN and found that there is still room for improvement. For example, they confirmed that an unnecessary mode called a droplet, which is drop-shaped noise, makes generated images less realistic and adds other minor unnatural features. Therefore, Karras et al. (2020b) proposed StyleGAN2, a revised StyleGAN architecture. In particular, the normalization method is redesigned, and residual connections are introduced to connect modules of different resolutions in the synthesis network and regularize the output from the synthesis network to increase the smoothness of the generator and improve the quality of the synthesized images.

To generate the synthesized damaged roof images, StyleGAN2 was trained on 714 images from the dataset mentioned in the previous section. The input data to the discriminator were augmented using adaptive discriminator augmentation (ADA) (Karras et al., 2020a), which adaptively augments the training data to avoid overfitting and mode collapse. The model's input batch size was eight, with 256 × 256 images. Both the discriminator and generator were optimized using Adam. In this study, the model was trained on a computer with an NVIDIA GeForce GTX 2080 8 GB GPU for five days. In this study, 250 images were generated for the primary investigation, and these images were purely generated from trained StyleGAN2, and they were not selected based on the quality or variety of the feature. Examples of real and generated images are shown in Figures 3 and 4.

Commented [A14]: Kindly double check if this addition improves readability.

Commented [A15]: **Tip:** Repeated noun error refers to the unnecessary repetition of the same noun in a series or list

For example,
Original: There were 12 and 18 members in the first *group* and second *group*, respectively.
Revision: There were 12 and 18 members in the first and second *groups*, respectively.

Figure 3. The examples of real damaged roof images

Figure 4. The example of generated synthesized damaged roof images

Division of roof images

In the next step, the generated synthesized damaged roof images were divided into roof segments. The division of roof images is necessary because the roof damage rate for issuing damage certificates is calculated using the following equation:

$$\text{Damage rate} = \sum_{i} \frac{\text{Area of roof segment}_i}{\text{Area of entire roof}} \times \text{Damage level of roof segment}_i \qquad (2)$$

As shown in Equation (2), the damage level must be calculated using the roof segment. To divide the roof image into roof segments, Fujita and Hatayama (2021) used a deep-learning-based instance segmentation model called mask R-CNN (He et al., 2018). Mask R-CNN achieved outstanding performance in the COCO challenge in 2016 for the tasks of instance, segmentation, bounding-box object detection, and person keypoint detection. The data to be divided is a rod of buildings in Japan, but the original mask R-CNN was trained with the COCO dataset, which mainly lacks Japanese-style roof images for segmentation. Therefore, Fujita and Hatayama (2021) fine-tuned the R-CNN pre-trained model using the dataset described in the previous section. This pre-trained model segmented the synthesized damaged roof images before post-processing divided them into roof segments; examples of these are shown in Figure 5. As a result, the divided synthesized damaged roof images augmented the original dataset used in Fujita and Hatayama (2021) with an additional 200–300 images (Table 1).

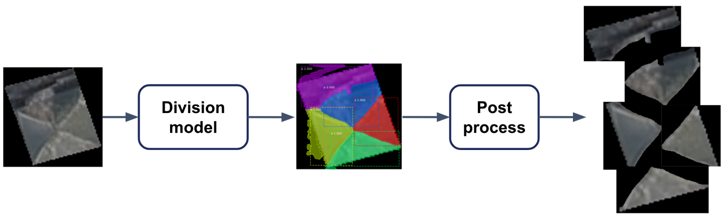

Figure 5. The Example of Dividing Synthesized damaged roof image in

Training of the classification model

Finally, the StyleGAN2-based data augmentation was verified using the damage rate classification model. For roof damage classification tasks, Fujita and Hatayama (2021) used a deep-learning-based classification model called ResNet (He et al., 2015). ResNet won the ILSVRC 2015 classification task and has become one of the most popular in computer vision models. Generally, increasing the number of layers improves the ability of deep learning models; however, there is a limit to the increase in the number of layers. Backpropagation is a technique for optimizing model parameters in deep learning models that propagate the computed loss of the last layer to the input layer. However, when the number of layers is large, the gradient of the computed loss vanishes in the middle of the hidden layers, causing the parameters to be incorrectly updated. ResNet addresses this problem by adding a residual block that bypasses the data from one layer to another (skip connections). The architecture of ResNet is widely used in many deep learning models.

The details of the training in (Fujita and Hatayama, 2021) are as follows. The model was trained using pre-trained ResNet50 and ResNet with 50 layers on the ImageNet dataset and fine-tuned for 500 epochs. The input size to the model was 256 × 256 with a batch size of eight, and 20% of the training data were used for validation. In this study, the model was trained on a computer with an NVIDIA GeForce GTX 2080 8 GB GPU for approximately 5 h.

Commented [A16]: Numbers below 11 must be written in words.

Commented [A17]: A space must not be added between a number and the SI unit.

RESULTS AND DISCUSSION

Table 2 summarizes the results of the classification evaluated with the test data as a confusion matrix, where the column of the table indicates the predicted results and the row of the table indicates the labels. The recall scores for each class were 71% for no damage, 44% for –25 damage, 25% for 25–50 damage, 23% for 50–75 damage, and 38% for 75–damage. Table 3 compares the recall score of Fujita and Hatayama (2021)'s study that used simple conventional augmentation, with the results of this study, which included StyleGAN2-based augmentation. The average recall score of the classification result trained with the proposed dataset improved by 3.9%. For the class with a small amount of raw data and a high ratio of simple augmentations, such as mirroring and rotation, in the original dataset, which are 25–50 damage, 50–75 damage, and 75–damage classes, the average reproduction rate increased from 17.3% to 28.7%, a 11.4% improvement. The results confirm that StyleGAN2-based data augmentation outperforms traditional data augmentation methods, particularly for classes with small sample size in the raw dataset. Although the effect of StyleGAN2-based data augmentation was verified, the proposed method has several limitations. The research conducted by Naito et al. (2020) shows that the regional dependency of the dataset affects the accuracy of the model; however, this is not considered in the proposed method. The proposed method generates synthesized images based on the existing post-disaster roof dataset, which cannot be extended to other regions. Data transformation is required to extend the damaged roof dataset to data from other regions. StyleGAN2 can generate new high-quality roof styles, but it is difficult to generate a specific type of roof that fits the target region. Additionally, a dataset of the target region is very likely to exist. Therefore, future research should consider alternative GAN models. Cycle-GAN (Zhu et al., 2020) is one possibility. In the targeting region, this model transforms the style between two unpaired datasets, such as the damaged roof and the non-disaster roof datasets.

Commented [A18]: **Tip:** A dummy subject doesn't convey any specific meaning other than simply serving as a grammatical filler. Sentences with a dummy subject most often begin with It or There.

For example,
Original: *There were* many people standing in line.
Revision: Many people were standing in line.

Original: *It is* your responsibility to open the gate.
Revised: You are responsible for opening the gate.

Table 3. Confusion Matrix of damage rate classification

	No Damage	–25 Damage	25–50 Damage	25–50 Damage	75– Damage
No Damage	173	45	7	3	12
–25 Damage	17	23	5	2	5
25-50 Damage	7	16	10	3	4
50–75 Damage	8	1	0	5	8
75– Damage	0	4	3	3	6

Table 4. Recall Scores

	No Damage	–25 Damage	25–50 Damage	25–50 Damage	75– Damage	Average
Fujita and Hatayama. 2021	75	56	13	14	25	36.3
Ours	71	44	**25**	**23**	**38**	**40.2**

CONCLUSION

This study investigated StyleGAN2-based data augmentation, and its validity was verified by the application of the roof damage rate classification task. The results show that for classes with small raw data sizes, the StyleGAN2-based data augmentation improves the recall score by 11.4% on average, and a high ratio of simple augmentations and overall recall score improved by 3.9%. Based on this result, the proposed augmentation method can be a breakthrough in machine learning applications in disaster prevention, where disaster data collection is difficult.

Currently, the dataset contains approximately 1000 images for each class; however, as discussed in previous sections, this is insufficient to achieve acceptable performance. Therefore, future work should focus on increasing the ratio of synthesized images in the dataset. Furthermore, an evaluation of the synthesized images is planned. The quality of the synthesized images is essential not only for improving the models' ability but also for ensuring their accountability. Accountability is important in the disaster prevention field because the decisions made by the model can be a matter of life; thus, users are very sensitive to the validity of the model's decisions.

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.com) for English language editing.

REFERENCES

Commented [A19]: I double-checked the references and found that they were all cited. Furthermore, the references are all consistent.

Anand, R., Mehrotra, K. G., Mohan, C. K., and Ranka, S. (1993). "An improved algorithm for neural network classification of imbalanced training sets." *IEEE Transactions on Neural Networks*, 4(6), 962–969.

Anand, V., and Miura, Y. (2021). "PreDisM: Pre-Disaster Modelling with CNN Ensembles for At-Risk Communities." *arXiv:2112.13465 [cs]*.

Antoniou, A., Storkey, A., and Edwards, H. (2018). "Data Augmentation Generative Adversarial Networks." *arXiv:1711.04340 [cs, stat]*.

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J., and Rueckert, D. (2018). "GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks." *arXiv:1810.10863 [cs]*.

Chaurasia, K., Kanse, S., Yewale, A., Singh, V. K., Sharma, B., and Dattu, B. R. (2019). "Predicting Damage to Buildings Caused by Earthquakes Using Machine Learning Techniques." *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, IEEE, Tiruchirappalli, India, 81–86.

DeVries, T., and Taylor, G. W. (2017). "Improved Regularization of Convolutional Neural Networks with Cutout." *arXiv:1708.04552 [cs]*.

Dolhansky, B., and Ferrer, C. C. (2017). "Eye In-Painting with Exemplar Generative Adversarial Networks." *arXiv:1712.03999 [cs, stat]*.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). "Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification." *arXiv:1801.02385 [cs]*.

Fujita, A., Sakurada, K., Imaizumi, T., Ito, R., Hikosaka, S., and Nakamura, R. (2017). "Damage detection from aerial images via convolutional neural networks." *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, IEEE, Nagoya, Japan, 5–8.

Fujita, S., and Hatayama, M. (2021). "Development of Automatic Method to Calculate Damage Rate of Roof by dividing Roof Image into Sections by Image Segmentation." Morioka, Japan.

Gao, Y., Kong, B., and Mosalam, K. M. (2019). "Deep leaf-bootstrapping generative adversarial network for structural image data augmentation." *Computer-Aided Civil and Infrastructure Engineering*, 34(9), 755–773.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). "Generative Adversarial Networks." *arXiv:1406.2661 [cs, stat]*.

Hamada, K., Tachibana, K., Li, T., Honda, H., and Uchida, Y. (2018). "Full-body High-resolution Anime

Generation with Progressive Structure-conditional Generative Adversarial Networks." *arXiv:1809.01890 [cs, stat]*.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2018). "Mask R-CNN." *arXiv:1703.06870 [cs]*.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Deep Residual Learning for Image Recognition." *arXiv:1512.03385 [cs]*.

Inoue, H. (2018). "Data Augmentation by Pairing Samples for Images Classification." *arXiv:1801.02929 [cs, stat]*.

Jabbar, A., Li, X., and Omar, B. (2020). "A Survey on Generative Adversarial Networks: Variants, Applications, and Training." *arXiv:2006.05132 [cs, eess]*.

Johnson, J. M., and Khoshgoftaar, T. M. (2019). "Survey on deep learning with class imbalance." *Journal of Big Data*, 6(1), 27.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2018). "Progressive Growing of GANs for Improved Quality, Stability, and Variation." *arXiv:1710.10196 [cs, stat]*.

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., and Aila, T. (2020a). "Training Generative Adversarial Networks with Limited Data." *arXiv:2006.06676 [cs, stat]*.

Karras, T., Laine, S., and Aila, T. (2019). "A Style-Based Generator Architecture for Generative Adversarial Networks." *arXiv:1812.04948 [cs, stat]*.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020b). "Analyzing and Improving the Image Quality of StyleGAN." *arXiv:1912.04958 [cs, eess, stat]*.

Kodali, N., Abernethy, J., Hays, J., and Kira, Z. (2017). "On Convergence and Stability of GANs." *arXiv:1705.07215 [cs]*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., Curran Associates, Inc.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. (2017). "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." *arXiv:1609.04802 [cs, stat]*.

Lee, K., Choi, C., Shin, D. H., and Kim, H. S. (2020). "Prediction of Heavy Rain Damage Using Deep Learning." *Water*, 12(7), 1942.

Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W., and Lin, L. (2018). "BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network." *Proceedings of the 26th ACM international conference on Multimedia*, ACM, Seoul Republic of Korea, 645–653.

Lin, K., Li, D., He, X., Zhang, Z., and Sun, M.-T. (2018). "Adversarial Ranking for Language Generation." *arXiv:1705.11001 [cs]*.

Liu, C., Ge, L., and Sepasgozar, S. M. E. (2021). "Post-Disaster Classification of Building Damage Using Transfer Learning." *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, Brussels, Belgium, 2194–2197.

Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., and Omata, H. (2021). "Generative adversarial network for road damage detection." *Computer-Aided Civil and Infrastructure Engineering*, 36(1), 47–60.

Miklosik, A., and Evans, N. (2020). "Impact of Big Data and Machine Learning on Digital Transformation in Marketing: A Literature Review." *IEEE Access*, 8, 101284–101292.

Mirza, M., and Osindero, S. (2014). "Conditional Generative Adversarial Nets." *arXiv:1411.1784 [cs, stat]*.

Miyamoto T., Asakawa T., Kubo H., Nomura Y., and Miyamori Y. (2020). "RESEARCH TRENDS IN MACHINE LEARNING FROM THE PERSPECTIVE OF DISASTER PREVENTION APPLICATIONS." Intelligence, Informatics and Infrastructure.

Miyamoto, T., and Yamamoto, Y. (2020). "Using Multimodal Learning Model for Earthquake Damage Detection Based on Optical Satellite Imagery and Structural Attributes." *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Waikoloa, HI, USA, 6623–6626.

Mogren, O. (2016). "C-RNN-GAN: Continuous recurrent neural networks with adversarial training."

arXiv:1611.09904 [cs].

Moradi, M., Madani, A., Karargyris, A., and Syeda-Mahmood, T. F. (2018). "Chest x-ray generation and data augmentation for cardiovascular abnormality classification." *Medical Imaging 2018: Image Processing*, E. D. Angelini and B. A. Landman, eds., SPIE, Houston, United States, 57.

Moreno-Barea, F. J., Strazzera, F., Jerez, J. M., Urda, D., and Franco, L. (2018). "Forward Noise Adjustment Scheme for Data Augmentation." *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, Bangalore, India, 728–734.

Naito, S., Tomozawa, H., Mori, Y., Monma, N., Nakamura, H., Fujiwara, H. (2020), "Development of the Deep Learning Based Damage Detection Model for Buildings Utilizing Aerial Photographs of Plural Earthquakes", *Journal of Japan Association for Earthquake Engineering*, 20(7), 177-216.

Radford, A., Metz, L., and Chintala, S. (2016). "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." *arXiv:1511.06434 [cs]*.

Roh, Y., Heo, G., and Whang, S. E. (2021). "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective." *IEEE Transactions on Knowledge and Data Engineering*, 33(4), 1328–1347.

Sandfort, V., Yan, K., Pickhardt, P. J., and Summers, R. M. (2019). "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks." *Scientific Reports*, 9(1), 16884.

Shorten, C., and Khoshgoftaar, T. M. (2019). "A survey on Image Data Augmentation for Deep Learning." *Journal of Big Data*, 6(1), 60.

Taylor, L., and Nitschke, G. (2017). "Improving Deep Learning using Generic Data Augmentation." *arXiv:1708.06020 [cs, stat]*.

Tilon, S., Nex, F., Kerle, N., and Vosselman, G. (2020). "Post-Disaster Building Damage Detection from Earth Observation Imagery Using Unsupervised and Transferable Anomaly Detecting Generative Adversarial Networks." *Remote Sensing*, 12(24), 4193.

Wickramaratne, S. D., and Mahmud, Md. S. (2021). "Conditional-GAN Based Data Augmentation for Deep Learning Task Classifier Improvement Using fNIRS Data." *Frontiers in Big Data*, 4, 659146.

Wu, H., Zheng, S., Zhang, J., and Huang, K. (2019). "GP-GAN: Towards Realistic High-Resolution Image Blending." *arXiv:1703.07195 [cs]*.

Xu, R., Xiao, Z., Huang, J., Zhang, Y., and Xiong, Z. (2021). "EDPN: Enhanced Deep Pyramid Network for Blurry Image Restoration." *arXiv:2105.04872 [cs]*.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2020). "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks." *arXiv:1703.10593 [cs]*.