ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol.12 (2022) No.1, pp. 1-12
DOI:10.33168/JSMS.2022.0101

An AI Based Algorithm for Restoration of Damaged Cultural Properties

ByongKwon Lee

School of Multimedia,   Seowon University, Korea

Sonic747@seowon.ac.kr

Abstract. Through the research of numerous researchers' artificial intelligence imitates human language and visual expression with good performance and imitates human style in voice and picture. This ability although dependent on the data for learning artificial intelligence is more objective and based on numerical data than humans. We applied it to the restoration of cultural assets made in the past through artificial intelligence neural networks and we applied a general CNN a little differently for the purpose of restoration. Cultural properties contain various backgrounds from the era when they were created and for this reason there are many complications and difficulties in restoration. If it is simply regarded as noise and recovered the result is dependent on the learned data. To solve this problem the CNN was separated into full and detailed and the association was learned together and the damaged part was repaired through a generative competition network (GAN) based on this neural network. We trained a neural network that extracts visual features on a Korean "Pagoda" (mostly produced under the influence of Buddhism) and conducted a study to repair the damaged part based on the trained neural network. The features of the tower were extracted through a CNN-based neural network and the damaged part was repaired through a Generative Adversarial Network (GAN) based on the extracted features. It is thought that our research will be actively used for the restoration of cultural assets as well as the restoration of archaeological records in the future

Keywords: Cultural Restoration, Artificial Intelligence Learning, Tower Learning Restoration, Hierarchical Repetition Learning, GAN Restoration

1. Introduction

This research focuses on restoring the old pagoda which are most Buddhist Leaning Tower in Korea. The tower is an intangible cultural property and has a 3D structure. However, based on the 2D image rather than the composition of 3D, the tower is recognized and describes the interesting part of the tower in detail, and is made in the range of restoring the damaged part. This research is based on the CNN(Convolutional Neural Network) (Abdelhak Belhi et al., 2020), which showed excellent results in recent image processing, and the GAN (Generative Adversarial Network), which showed good results in the field of image in painting. And we researched for CNN (Capsule Neural Network) and DCGAN (Depth-Convolution Generative Adversarial Network) were reviewed together and organized hierarchically. Cultural properties have different characteristics in periods and regions and contain various information of the period. Currently, the restoration of cultural properties utilizes the historical and regional characteristics of the remaining records, or the damaged or lost cultural properties through comparison with similar buildings (Gregory Koch et al., 2015). However, it is difficult to maintain objectivity because the restoration of cultural properties depends on the experience and knowledge of experts, and it takes a lot of time to confirm that the restoration results are correct because the cultural properties have the characteristics of the times(Krizhevsky, A et al., 2012).

2. Proposed method

2.1. Data Review - Characteristics of the tower

In order to learn artificial intelligence, we need more detailed classification of the data to be trained. Fortunately, because Korean pagodas can be classified according to their material, purpose, and origin, the shape of the excavated pagoda can be inferred even a little (J. Dai et al., 2015). There are many elements that categorize the tower, and we have summarized three important visual parts of them.

① Stone pagoda: A pagoda made of stone to enshrine the Buddha's sari. In Korea, there are many high-quality granites, so stone pagodas began to be built early and occupy most of the pagoda.

② Front tower: a tower made of bricks. In China, wooden pagoda was originally the standard, but from the North and South Korean dynasty, the eaves of wooden constructions and pagodas that imitated the headgear were popular.

③ Wooden pagoda: A pagoda made of wood. It is rare in India, but it is a form of a pagoda that was popular in China.

④ Mojeon Stone Pagoda: A pagoda made of stone and imitating the previous pagoda. Its shape can be divided into two types: the stone was cut like a brick and built into a tower shape, and the surface was processed and built like a tower while following the basic form of a general stone tower.

⑤ Dabo Pagoda, the pagoda that enshrines Dabo Buddha one of the past Buddhas. It is also called Dabo Buddha Pagoda or Chilbo Pagoda. The Buddhist temple and Dabo Buddha were placed together in the pagoda by the beophwagyeong gyeongbo

pagoda.
⑥ Bohyupin tower: It refers to a tower in which Bohyup Indala Nigyeong is inserted. The base and the tower are all rectangular in shape, with a leaf-shaped ornament attached to the top corner and an upper ring in the middle. In addition, it is common to engrave a panja or Buddha image on the four sides of the pagoda.
⑦ Stupa: A tower to enshrine the Buddha's sari. It originated from the fact that each country built a stupa by dividing the sari obtained by cremating the remains after the Buddha's death and distributing it into 8 parts.
⑧ Sasajaseok Pagoda: The pagoda priest follows the general stone pagoda style, but it refers to a form in which lions are placed at each of the four corners of the upper air base and a standing or left statue is placed in the center.
⑨ Cheongseok Pagoda: A pagoda made of a peculiar blue-colored stone called slate-am. The basic form is the same as the general type stone pagoda, but all are small towers because the stone itself is thin and small. Due to these material properties, the base end of this stone pagoda is made of granite(K.He et al., 2014)(A. KrizhevSky et al., 2012).

2.2. Classification type and the times

When categorized by shape, it is largely divided into general pagoda, deformed pagoda or special pagoda (special type) transformation and special type. According to the classification by period classification, it is divided into Goguryeo Dynasty Pagoda, Baekje Dynasty Pagoda, Go Silla Dynasty Pagoda, Unified Silla Dynasty Pagoda, Goryeo Dynasty Pagoda, Joseon Dynasty Pagoda, etc. it can be divided into the predecessor (early period) · the early period, the middle period, the late period and the end period [Figure 1]. In Figure 1, the picture on the left is a restored stone pagoda in the past, and the picture on the right is a restored picture of the present. In conclusion, it was confirmed that the restoration technique differs in the restoration method and knowledge according to the era. So, in the study, we used an artificial intelligence algorithm to lock in consistency regardless of the point in time of restoration.

Fig 1: Restoration of Iksan Mireuksaji stone pagoda

2.3. Difficulty of restoration

Many attempts have been made through experts to restore the tower. The recently restored “Iksan Mireuksaji Stone Pagoda” is the result of restoration by analogizing the technology of that period from the lower part according to the method of manufacturing the pagoda (Fig. 1).

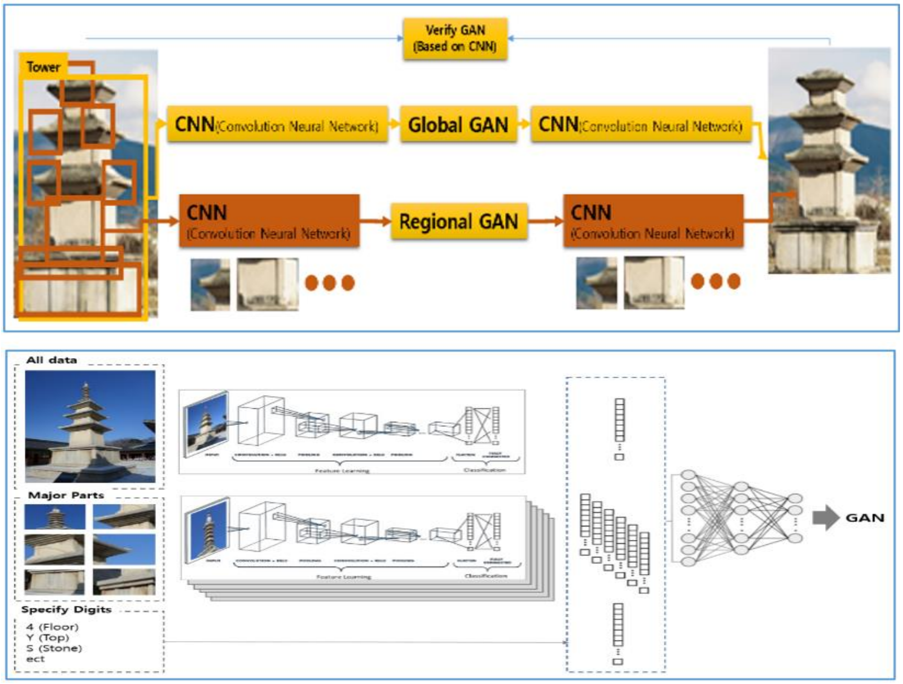

Fig 2: Layout for CNN, GAN and Extended CNN

However, despite this recovery, it is very difficult to answer various questions from experts. The reason is that the stone pagoda that existed in that era and the influence of political and domestic and foreign conditions are included. Therefore, we excluded these conditions to restore the tower, and experimented based on the data that had been restored or the originals existed. In Fig 2, We trained intact towers through a common CNN (J.Long,E et al., 2015)(R.Grishick et al., 2014). However, the characteristic parts that make up the tower were studied together, and the data representing the tower were also learned. However, it was difficult to secure enough data that could be obtained for research, so we propose a neural network such as Siamese Neural Networks, which is extended from a general CNN (P.Agrawal et al., 2014). The neural network was trained by configuring the overall outline and detailed parts of the tower, and the features of the tower classification as a single network. The following shows the overall network for feature extraction. Details on training will be covered in the main text of this paper.

2.4. GAN (Generative adversarial networks)

GAN (Generative adversarial networks) is a technique for generating images and has been researched and published very quickly since it first appeared in 2014. GAN can not only create images (K. Simonyan et al., 2015), but also improve style transfer and application and image recognition performance for images. GAN is a generative model that operates in reverse order the discriminative model that finds the correct answer (y) when there is an input image (x) for data processing (classification, detection, segmentation) in CNN (C. Szegedy et al., 2014). The generative model was studied in the form of a discriminative model. As representative models, method studies of GMM (Gaussian Mixture Model) and HMM (Hidden Markov Model) have been conducted. In addition, the GAN learning model performs learning for image generation as adversarial learning (D. Erhan et al., 2014). The operation method of the GAN learning model works as follows when there are two models. Model A finds the vulnerabilities of learned model B and learns disturbance, and model B learns in the direction of compensating the discovered vulnerabilities.

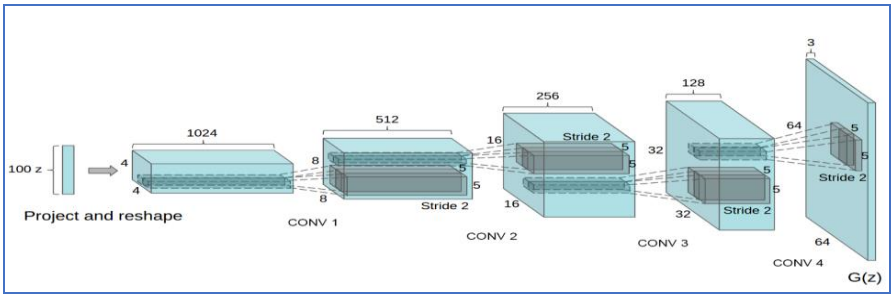

Fig 3: Construction of Deep Convolution Generator for GAN.

Adversarial learning used in GAN proceeds similarly, with a score between the two and one (G) to minimize the score and one (D) to the maximum. Competitive scores are expressed by the following formula. [minGmaxDV(D,G) = Ex ~ pdata(x)[log(D(x))] + Ez ~pz(z)[1-log(D(G(z)))]] However, rather than using the full connection used in GAN, deep convolutional adversarial networks (DCGAN) suitable for 2D images are used. In general image processing, it is difficult to measure the reliability of comprehensive GAN, this is because it is difficult to verify the full connection, but it is more advantageous for image generation, including correlation with space. In Figure. 3, the image generator for restoration has the following structure (Ian Goodfellow et al., 2014).

2.5. Autoencoder

Assuming that there is a structure like the above, it is to make the output appear as identical as possible to the input. Autoencoder differs in the number of inputs and outputs and the number of nodes of the hidden layer in the middle. That means that the dimension is reduced in the hidden layer, and if the output represents the input well enough, the reduction will be a very meaningful reduction. In addition, it is possible to design a problem in which noise is added to the input data to be input and encoded, and the original data without noise is restored through a decoder. Through this problem setting, the goal is not to simply output the input as an output, but to make a good intermediate state of representing in order to convert the input into output. Autoencoder belongs to the following learning categories. It brings good results in learning the image of the tower we want to do {Unsupervised Learning}, {Representation Learning}, {Dimensionality Reduction}, {Generative Model Learning} In general, autoencoder has a symmetrical structure of encoder and decoder. we can create the desired result image using the latent vector created/extracted above. In terms of this generation, the decoder can be a generator. The representative generative model is GAN but there is also VAE (Yujia Li et al., 2020).

3. Experiment

3.1. Optimized GAN and CNN algorithm for the restoration

As mentioned in the introduction, cultural properties have characteristics of the times, and it takes a lot of time to confirm that the restoration results are correct. To solve this problem, we apply artificial neural networks to ensure reliability and objectivity.

3.2. CNN algorithm for feature point extraction

First, a convolutional neural network (CNN) was used as a neural network to distinguish feature points on cultural properties. CNN is an algorithm that shows good results for object detection and object recognition, which has the advantage of recognizing cultural properties and classifying characteristics through images. However, if the CNN algorithm is used without considering the diversity of the tower (cultural property), there is a problem that learning takes place without including the characteristics of the times. Fig 3 shows the change of the tower similar to the times in Korean history. As a result, although there are other characteristics according to the flow of the times, the flow by times is very similar, so if you find and classify the feature points, you can learn about the tower. In addition, Fig 2 shows the Dabotap of Bulguksa Temple in Gyeongju has a very unique structure in the times, so additional information about the Dabotap was added to the CNN algorithm to study.

The CNN was constructed so that additional information about the Dabotap could be divided into layers and connected in layers. As a result, even though they were classified as completely different towers through hierarchical neural network construction, they were able to find relevance through dependent CNN learning.

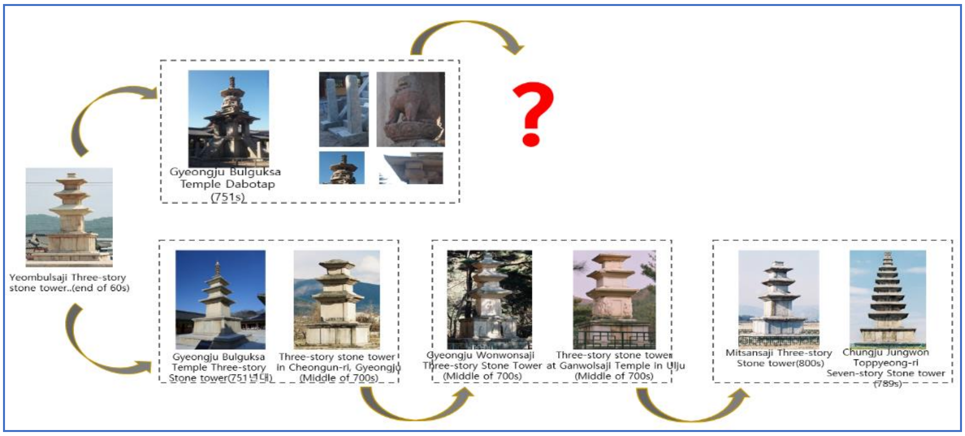

Fig 4: Classification of Cultural Properties (Towers) of the Unified Silla Period of Korea

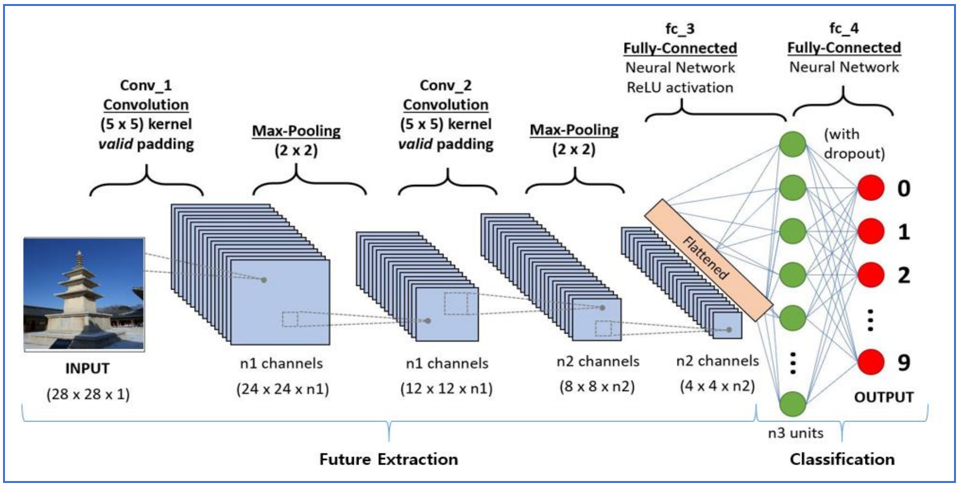

Fig 5: Learning procedure using CNN

However, the proposed CNN learning model has a lower accuracy because the size of the upper and lower parts of the tower is different in the case of a pyramid-shaped tower. To solve this problem, applying CNN's polling can reduce the learning time and create a robust neural network for changing the position of image elements. The proposed Max Pooling technique sends only the maximum value of the inference result of the surrounding area to the upper layer. The feature search area (Feature Map) is usually reduced to 1/4, and the processing capacity for feature extraction and inference in the upper layer is greatly reduced as shown in Figure 4. In addition, there is an advantage of obtaining an invariance that is advantageous for classifying characteristics. For example, if features are extracted differently depending on the direction or angle of the tower, it cannot be properly recognized in the upper layer. However, the Max Pulling technique allows you to recognize the features that make up the tower regardless of its location. The advantage of Max Pooling is that it extracts excellent results in recognizing the tower, but the detected location information disappears in the field of tower. In other words, the Max Pooling method designed to speed up the search has a vulnerability in that the detected location and spatial information disappear. In addition, the Max Pooling method based on CNN has an advantage of recognizing an object regardless of its location, but generates an error of recognizing an object having different directions and ratios for the same object as the training data. The reason for this problem is that in restoring cultural properties, it is necessary to maintain the relative position of the structure and restore the original shape. Cases of failure of a convolutional neural network is a failure example of Max Pooling based on CNN and recognizes the face the same regardless of its position. if CNN-based Max Pooling technique is applied to the tower, it is possible to recognize objects placed in different locations as the same tower. The extended CNN model proposed by us is to classify according to the location of objects by including additional information in a specific layer. As a result, if there are towers with similar features, they will be classified into different towers according to the location of the object.

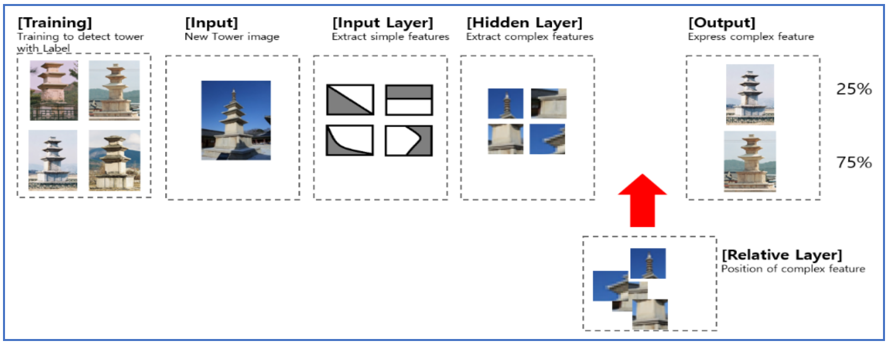

Fig 6: Relative position layer insertion

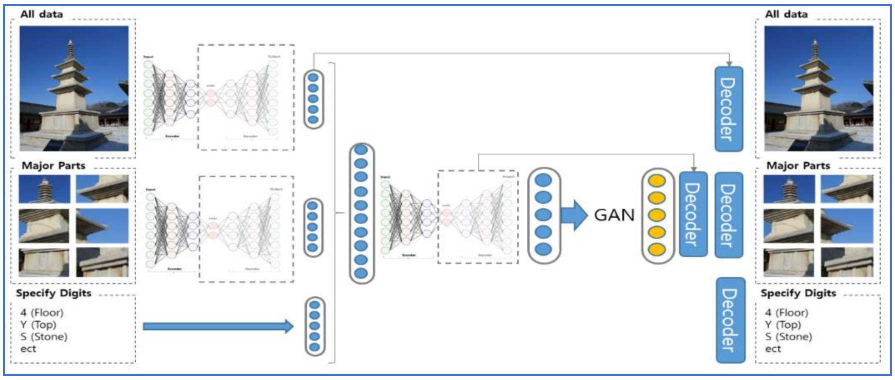

Fig 7: Proposed neural network based on CNN & GAN

The towers of Korea's Unified Silla period (see Fig. 1) have a similar shape, the order and location of objects on each floor are also important. Fig 5. shows the relative location layer added in the CNN learning process. The Cultural properties having the same characteristics as the relative position values of the tower structure. Also, it can be recognized as different towers depending on the location. The proposed technique was conducted in 2017 by Geoffrey Hinton et al. It is a capsule neural network proposed by the professor. The capsule neural network technique recognizes the shape of the tower as two different towers when there are two different towers. The capsule neural network structure is a unit element composed of several neural network layers, and the "squashing function" is applied to the entire vector by dynamic routing between nested layers. In the study, the MNIST data set was trained based on the designed capsule neural network to reduce the error rate compared to CNN by 45%. It is also more effective against white box adversarial attacks. We put these advantages together and devised the following neural network as shown in Fig 6.

4. Discussion

In order to reinforce the neural network for the lost part, we divided it into all data and major parts. The parts that we want to recover from the damaged tower through this study are the overall outline and major restoration parts of the tower. It is also a descriptive feature that can represent the tower. Step 1(feature training): The given data may not be a complete image, and when acquired, various data losses occur due to outdoor conditions. Therefore, Denoising Autoencoder was used to recover the lost part. The stacked neural network is trained from hidden layer 1 inward. The reason for separating and learning in this way is to save the learning time and to lighten the neural network by classifying the features we want to recover. Step 2(correlation training): Features extracted through auto-encoder are re-learned through general auto-encoder to extract related features. The reason why each learned data is not used directly as input to the GAN is because of the corrupted data to be restored. The second auto-encoder receives inputs of the three merged data and learns its function and association. With stacked encoders, neural networks can be strongly trained on the lost part, and the part to be recovered can be created through the overall characteristics. Step 3(GAN training): The feature vectors created in this way are learned and generated through GAN. In general, to minimize the difficulty of learning GAN and failure to apply neural networks due to lack of data, the scope of use of GAN has been minimized. Also, by learning only the characteristics and associations of the intact tower, it can be sufficiently generated even if a damaged tower is entered later. GAN will competitively generate features extracted through the learned autoencoder. By limiting the role of GAN, training time can be shortened and problems with little data can be compensated. Step 4(generation): Vectors generated through GAN are played through decoders learned by auto encoder. Instead of relying solely on the GAN for image generation, you can use an existing trained neural network to generate undiscovered or lost parts. As with the input, based on the generated data, the original full image, partial image, and characteristic data of the tower are restored together. The neural network, which is strongly trained in both directions, recovers the entire image and the partial image with only the part excavated from the cultural property restoration. Step 5(model verification): It is difficult to find a way to evaluate the performance of GAN. In this case, it was evaluated using the Inception Score (IS) based on Google's Inception model, and the performance evaluation index for the neural network was Fréchet Inception Distance (FID). The presented indicators were measured for performance using the Inception network. The Fréchet Inception Distance (FID) uses the Inception network Inception V3, which measures the distance in the feature space of real data and fake data. Equation 2 extracts the features of the photo of the tower and the photo generated by the GAN, calculates the mean and covariance (mr, Cr), (mf, Cf) of the two sets of features and calculates the distance by value.

FID2=||mf-mr||22+Tr(Cf+Cr-2(CfCr)1/2) (1)

As a result of the test, it was confirmed that the trend of changes in the Loss of the two models was different during the learning. In the proposed method, it was confirmed that the GAN model method was not well trained in the initial stage because the generator was not fully trained. It was found that the learning batch size of the FID (Fréchet Inception Distance) was small, so it did not decrease more than a certain amount. In addition, the FID of data sampled at a sufficiently large size has been reduced. In conclusion, the number of FIDs according to learning is 00.00%, and considering that it is generally less than 10.00% when learning is good, it has not yet generated enough data to recover cultural properties. The reason is that the amount of sample data prepared in the study is insufficient Fig 8.

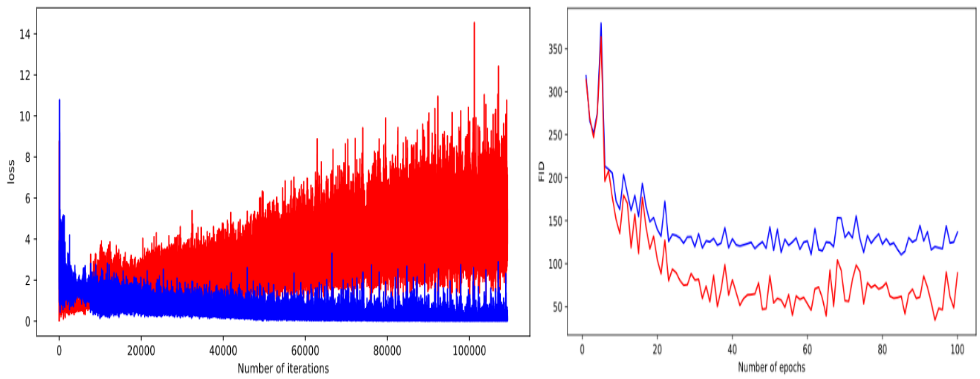

Fig 8: Fréchet Inception Distance (FID) values of Generator and Discriminator

5. Conclusion

Through this research in order to make the special task of restoring cultural properties more efficiently. We were able to maintain the overall structural characteristics of the tower through the capsule neural network based on the existing CNN and GAN. It is composed of neural networks so that each feature can be learned and implemented more naturally. Through this kind of combining neural networks, we can divide and clarify the role of each neural network in restoring cultural properties. CNN, a dependent neural network, can only focus on recognition and classification. Therefore, similar results with high performance can be obtained even if the dependent neural network uses Fast R-CNN or YOLO. However, here is not enough data that can be learned by the characteristics of cultural properties. We trained this complex neural network by producing data (Photos) for various angles and characteristics of cultural properties that were preserved completely by age. GAN-based extended DCGAN was used for the restoration of damaged cultural properties. Although GAN is a competitive training algorithm, it has been proven that sufficient training can produce results equal to or greater than humans. The data generated through DCGAN (originally created by Fake data) is similar to the original data, but because of the lack of detail, it is possible to recover one tower through one more dependent DCGAN for the characteristic aspect of the separately subordinate tower (originally Fake creation result). Therefore, the actual constructed artificial neural network is a capsule-generating hostile network based on CNN and DCGAN (Capsule GAN). In this research, we implemented a learning algorithm to recover a tower that is dependent on natural conditions such as buildings. However, many cultural properties are not subject to this restriction such as paintings, ornaments, and tiles. If this study is extended through RNN in the future, it may be possible to create various cultural properties by learning changes in the time sequence.

References

Abdelhak Belhi, Abdelaziz Bouras, Abdulaziz Khalid Al-Ali, Sebti Foufou (2020). A machine learning framework for enhancing digital experiences in cultural heritage. *Journal of Enterprise Information Management*, Vol. ahead-of-print No. ahead-of-print. DOI: https://doi.org/10.1108/JEIM-02-2020-0059

Gregory Koch, Richard Zemel, Ruslan Salakhutdinov (2015). Siamese Neural Networks for One-shot Image Recognition, *ICML deep learning workshop*, WEB: https://www.bibsonomy.org/bibtex/26f83b8c4cf316e77e6f6ce1e97411b30/bsc

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 1, 1097–1105. DOI: https://dl.acm.org/doi/10.1145/3065386

J. Dai, K. He, and J. Sun (2015). Convolutional feature masking for joint object and stuff segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*. DOI: https://doi.org/10.1109/CVPR.2015.7299025

K. He, X. Zhang, S. Ren, and J. Sun (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. DOI: https://arxiv.org/abs/1406.4729

J. Long, E. Shelhamer, and T. Darrell (2015). Fully convolutional networks for semantic segmentation. *In CVPR*. DOI: https://arxiv.org/abs/1411.4038

R. Girshick, J. Donahue, T. Darrell, and J. Malik (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *European Conference on Computer Vision*. DOI: https://arxiv.org/abs/1407.1610

P. Agrawal, R. Girshick, and J. Malik (2014). Analyzing the performance of multilayer neural networks for object recognition. *European Conference on Computer Vision*. DOI: https://doi.org/10.1007/978-3-319-10584-0_22

K. Simonyan and A. Zisserman (2015). Very deep convolutional networks for large-scale image recognition. DOI: https://arxiv.org/abs/1409.1556

C. Szegedy, A. Toshev, and D. Erhan (2013). Deep neural networks for object detection. *ERCVAD Workshop*. DOI: https://arxiv.org/abs/2108.08166

D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov (2014). Scalable object detection using deep neural networks. DOI: https://arxiv.org/abs/1312.2249

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio (2014). Generative adversarial nets. *In NIPS*. DOI: https://arxiv.org/abs/1406.2661