Restoration of Hand-Drawn Architectural Drawings using Latent Space Mapping with Degradation Generator

Nakkwan Choi[1], Seungjae Lee[2], Yongsik Lee[2], Seungjoon Yang[1*]
[1]Dept. of Electrical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Korea
[2]Electronics and Telecommunications Research Institute, Daejeon, Korea
{cvvc1997, syang}@unist.ac.kr, {seungjlee, ys.lee}@etri.re.kr

Abstract

This work presents the restoration of drawings of wooden built heritage. Hand-drawn drawings contain the most important original information but are often severely degraded over time. A novel restoration method based on the vector quantized variational autoencoders is presented. Latent space representations of drawings and noise are learned, which are used to map noisy drawings to clean drawings for restoration and to generate authentic noisy drawings for data augmentation. The proposed method is applied to the drawings archived in the Cultural Heritage Administration. Restored drawings show significant quality improvement and allow more accurate interpretations of information.

1. Introduction

Cultural heritage is a valuable asset of humanity that requires our efforts to preserve archaeological, historical, cultural, and technological values. In particular, traditional wooden buildings are vulnerable to deformation, earthquakes, and fires. We continuously collect and manage architectural drawings, photos, and 3D scan data of individual buildings for preservation and restoration. Among them, architectural drawings in the past contain initial information on traditional wooden builds, and their value is the most significant. However, many archived drawings in the form of scanned images are already degraded over time, making it difficult to interpret information due to noise and damage. There is a need to restore aged drawings to facilitate information interpretation. This work reports an effort to restore aged drawings of wooden built heritage archived by the Cultural Heritage Administration.

Aged drawings in the archive often show compound degradation with faded and deteriorated lines, smeared and blurred complex parts, and the background in faded color with smudged leakages from adjacent drawings. Restoration requires removing noise in the background, linking broken lines, and clearing up complex parts. A learning-based restoration method would be a good match since it can model such complicated degradation. However, while clean and noisy drawings are abundant, clean and noisy pairs of the same drawings are scarce. Modeling and restoration by supervised learning would be problematic. Synthetically generated clean and degraded image pairs can be used. [10, 16, 21–23] But the domain gap between the synthetically degraded drawings and the actual aged drawings may cause inferior restoration performance.

In this work, we proposed a vector quantized variational autoencoder [13] (VQ-VAE) based restoration method to restore aged hand-drawn architectural drawings. The proposed method consists of two stages. In the first stage, a VQ-VAE is trained to learn accurate latent space representations of clean drawings using a large set of clean drawings. In the second stage, a mapping of latent space variables of noisy drawings to those of clean drawings as well as the generator that produces realistic degraded drawings is learned. Degradation generator is trained to generate a noisy drawing with the residual mapping error as an input. The latent space mapping is learned using a set of drawing pairs, for which we use the outputs of the degradation generator as data augmentation. Noisy drawings generated by the degradation generator provide authentic variations of degradations clean drawings can suffer. Hence, the latent space mapping from noisy to clean drawings is more accurately learned, and the detrimental effect on the restoration performance caused by the domain gap can be mitigated.

The proposed method is applied to restore archived aged architectural drawings of traditional wooden buildings. Restoration performance was compared to other methods developed for heavy degradations of drawings and photographs. The proposed methods reported significant improvement in both quantitative measures and qualitative evaluations. The performance gain is the most apparent with actual aged drawings from the archives. The proposed degradation generator produced more authentic degraded

drawings than other generative methods. The data augmentation with the degradation generator during the training allowed accurate latent space mapping could be learned for the restoration.

The following can summarize the novelty of the proposed method. i) an effective VQ-VAE based restoration method for aged architectural drawings was proposed. ii) significant improvement in restoration performance was achieved in both quantitative measures and subjective evaluations compared to existing learning based restoration methods, and iii) a degradation generator, which generates more realistic degradation, was developed for data augmentation to generalize the model.

2. Related Work

Studies on the digitization of technical drawings address how to restore the degraded line structures in the drawings [2,6,8]. However, most digitization techniques perform poorly with noisy drawings that suffer severe degradation. Recently, studies have been on learning the representations of images that suffer domain-specific degradations. In [1,4], Unet-based denoising was utilized to improve vectorization of technical drawings. In [12, 15, 17, 18], representations of hand-line drawings were learned to detect, clean up, and complete line structures in sketches. In [16], the restoration of aged drawings archived in poor conditions was presented to restore deteriorated old sketches by Leonardo da Vinci. Two sub-networks, one for extracting the line segments and the other for completing the line drawing, were used to restore degraded sketches.

Domain-specific representation of data can be learned with deep networks. In [19], two VAEs were prepared, one for the representation of degraded images and the other for the representation of clean images. Then, a deep network that maps the latent space of degraded images to that of clean images is learned for restoration. In [20], two generators, one from clean to degraded and the other from degraded to clean images, are learned in an adverserial problem. In [22], multi-stage architecture is used to learn progressively how an image is degraded for restoration. Leaning latent space representation via VAE allows the generation of realistic images. In [13, 14], vector quantization is applied to discretize the latent variable, which results in high quality images and videos. These works are mainly focused on the restoration of heavily degraded photographic images. However, their approaches can be adopted for the restoration of aged architectural drawings.

This work addresses the similar problem in [16] of restoring aged drawings that suffer heavy degradation. The overall architecture is based on VQ-VAEs. We use one VQ-VAE to learn the representation of clean drawings. We use VQVAE-based model to learn a mapping of the latent variables of a noisy drawing to those of a clean drawing for restoration and also to generate realistic aged drawings for data augmentation in training.

3. Method

The proposed method consists of two parts. First, latent space representations of clean drawings are learned using a VQ-VAE. Second, a mapping of latent space variables from noisy drawings to clean drawings is learned. A degradation generator produces noisy versions of clean drawings for augmentation. The latent space mapping is learned using synthetically degraded drawings and the degradation generator to augment training data. After the training, a noisy drawing is restored by mapping the latent variables from noisy to clean space, which the VQ-VAE decodes. The schematic of the proposed method is shown in Fig. 1.

3.1. Latent Space Representation with VQ-VAE

In the first stage, the encoder E_c, generator G_c, and the codebook $\mathcal{C}$ of the vector quantizer are trained with a set of clean drawings. The encoder E_c takes a clean input $\mathbf{x}_c$ and produces a latent variable $\mathbf{z}_c$

$$\mathbf{z}_c = E_c(\mathbf{x}_c), \tag{1}$$

which is quantized and embedded by

$$k_c = Q(\mathbf{z}_c; \mathcal{C}), \tag{2}$$
$$\hat{\mathbf{z}}_c = \mathbf{e}_c(k_c; \mathcal{C}), \tag{3}$$

where Q is the quantizer, $\mathcal{C}$ is a codebook, k_c is the quantizataion index, and $\mathbf{e}_c$ is the codeword embedding, respectively. The generator G_c takes the embedded latent variable $\hat{\mathbf{z}}_c$ and produces the clean input back in the autoregressive setting,

$$\hat{\mathbf{x}}_c = G_c(\hat{\mathbf{z}}_c). \tag{4}$$

VQ-VAE for the clean drawings is trained using a set of clean drawings as a training set. Reconstruction loss

$$\mathcal{L}_{\text{recon}} = \|\mathbf{x}_c - \hat{\mathbf{x}}_c\|_2^2 \tag{5}$$

and the adversarial loss [5]

$$\mathcal{L}_{\text{adv}} = \min_{G_c} \max_{D_c} \mathbb{E}[D_c(\mathbf{x}_c)^2] + \mathbb{E}[(1 - D_c(G_c(\mathbf{z}_c)))^2] \tag{6}$$

are used to train the encoder E_c and the generator G_c with discriminator D_c. For training tokenization drawings as the codebook $\mathcal{C}$, the vector-quantization loss is used:

$$\mathcal{L}_{\text{VQ}} = \|\text{sg}(\mathbf{e}_c(k_c; \mathcal{C})) - \mathbf{z}_c\|_2^2 + \|\mathbf{e}_c(k_c; \mathcal{C}) - \text{sg}(\mathbf{z}_c)\|_2^2, \tag{7}$$

where $\mathbf{sg}$ is the stop gradient operation. The overall loss is given by

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}} + \lambda_{\text{VQ}}\mathcal{L}_{\text{VQ}} \tag{8}$$

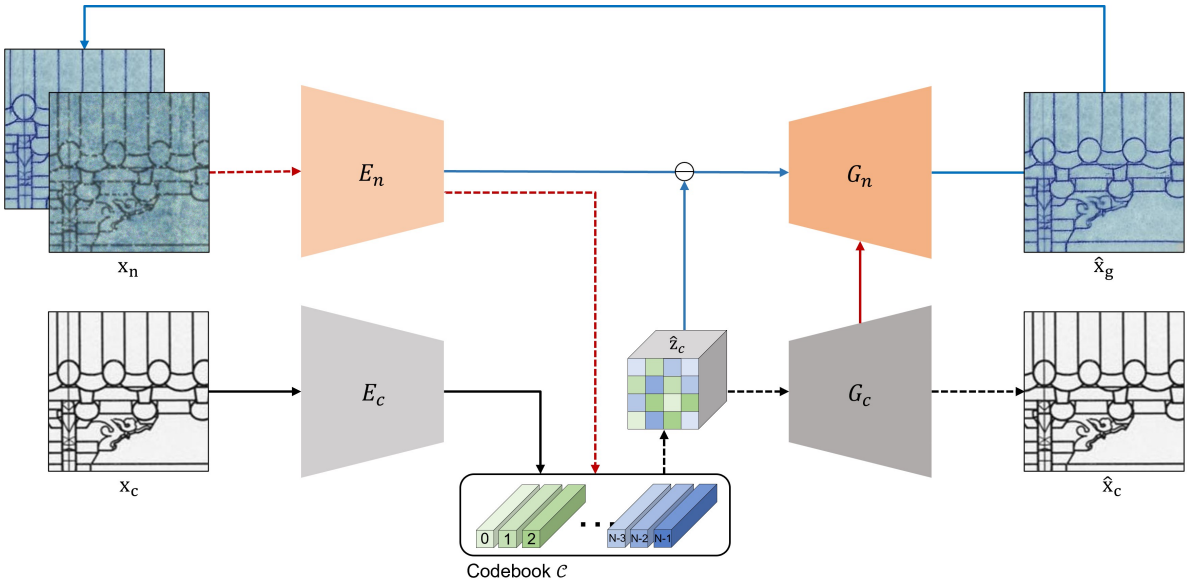


Figure 1. Architecture of proposed model. Black line: learning latent space representation with VQ-VAE. Red line: learning latent space mapping for restoration of noisy drawings. Blue line: degradation generator for data augmentation. Inference is through the path in dashed lines.

where λ_{adv} and λ_{VQ} weight the importance of the individual loss functions.

A large amount of clean architectural drawings are available for the training of the VQ-VAE. Hence, an accurate VQ-VAE for clean drawings can be obtained from the first stage of the training.

3.2. Latent Space Mapping for Restoration

In the second stage, the encoder E_n is trained so that it maps a noisy drawing input to the latent variables, which are quantized and decoded to be a clean drawing. The encoder E_n takes a noisy drawing $\mathbf{x}_n$ and produces a latent variable $\mathbf{z}_n$, which are quantized using the codebook $\mathcal{C}$.

$$\mathbf{z}_n = E_n(\mathbf{x}_n), \tag{9}$$
$$k_n = Q(\mathbf{z}_n; \mathcal{C}), \tag{10}$$
$$\hat{\mathbf{z}}_n = \mathbf{e}_c(k_n; \mathcal{C}). \tag{11}$$

The generator G_c restores a clean drawing $\hat{\mathbf{x}}_c$

$$\hat{\mathbf{x}}_c = G_c(\hat{\mathbf{z}}_n). \tag{12}$$

The encoder E_n is trained with a set of clean and noisy drawing pairs so that the latent space of noisy drawings corresponding to the clean drawings are mapped to the latent space of clean drawings. The VQ-VAE, i.e., the encoder E_c, generator G_c, and the codebook $\mathcal{C}$, trained in the first stage is frozen in the second stage. The encoders E_c and E_n map clean and noisy drawings $\mathbf{x}_c$ and $\mathbf{x}_n$ to the latent variables $\mathbf{z}_c$ and $\mathbf{z}_n$, respectively. The training of the encoder E_n is with the mapping loss

$$\mathcal{L}_{\text{map}} = \|\mathbf{z}_c - \mathbf{z}_n\|_2^2, \tag{13}$$

and the refinement loss

$$\mathcal{L}_{\text{ref}} = \|\hat{\mathbf{z}}_c - \hat{\mathbf{z}}_n\|. \tag{14}$$

The overall loss for mapping process is given by

$$\mathcal{L} = \mathcal{L}_{\text{map}} + \lambda_{\text{ref}} \mathcal{L}_{\text{ref}} \tag{15}$$

where λ_{ref} weighs the two looses.

Training of the latent space mapping by the encoder E_n requires a set of clean and noisy drawing pairs. In reality, there is no clean drawing corresponding to the noisy drawing. Therefore, we use the synthesis drawing by adding noise to the clean drawing. But complex patterns of actual aged drawings are not easy to synthesize, which degrades the restoration performance. The method for reducing the domain gap between the synthesized drawing and the actual aged drawing is described in the following sections.

3.3. Degradation Generator for Data Augmentation

The generator G_n produces noisy drawings from the residual mapping errors in the latent space. The input to the generator G_n is given by

$$\mathbf{z}_g = \lambda_e \mathbf{z}_e + \lambda_n \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{16}$$

where $\mathbf{z}_e$ is the residual error between the coded and embedded latent variables of the clean and noisy drawings of a

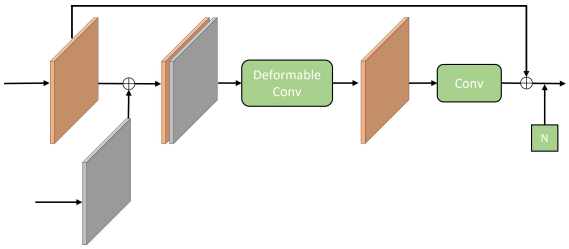


Figure 2. Degradation generator modifications. Intermediate activations of generator G_c is fed to generator G_n, which is followed by deformable convolutions. N stands for noise input for stochastic variation of degradation.

pair

$$\mathbf{z}_e = \mathbf{z}_n - \hat{\mathbf{z}}_c, \tag{17}$$

and $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is the normal distribution. The parameters λ_e and λ_n weight the mix of the residual error and the Gaussian noises.

The generators G_n and the generator G_c are based on the same architectures. There are two modifications for the generator G_n for the accurate modeling of the degradation. First, the intermediate activation of the generator G_c are fed to the generator G_n, which are concatenated to the intermediate activation of the generator G_n at the same levels. The output of the generator is

$$\hat{\mathbf{x}}_g = G_n(\mathbf{z}_g; G_c, \hat{\mathbf{z}}_c). \tag{18}$$

Second, the convolution operations are replaced by the deformable convolutions [3]. The deformable convolution allows informative line features of a drawing to be involved in the convolution operations and prevents cases where there are only blank spaces in receptive fields. By the two modifications, the information of a clean drawing is provided to the generator G_n, so that it can produce a noisy drawing by degrading the clean drawing. The modification of the generator G_n is shown in Fig. 2.

Generator G_n for the noisy drawings is trained using a set of clean and noisy drawing pairs. The reconstruction loss

$$\mathcal{L}_{\text{recon}} = \|\mathbf{x}_n - \hat{\mathbf{x}}_g\|_2^2 \tag{19}$$

is used for the training with the adversarial loss for the discriminator D_n:

$$\mathcal{L}_{\text{adv}} = \min_{G_n} \max_{D_n} \mathbb{E}[D_n(\mathbf{x}_r)^2] + \mathbb{E}[(1 - D_n(G_n(\hat{\mathbf{z}}_n)))^2] \tag{20}$$

where $\mathbf{x}_r$ is the actual aged drawing. The discriminator D_n is trained to distinguish noisy drawings generated by the degradation generator from actual aged drawings. The overall loss is given by

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}}. \tag{21}$$

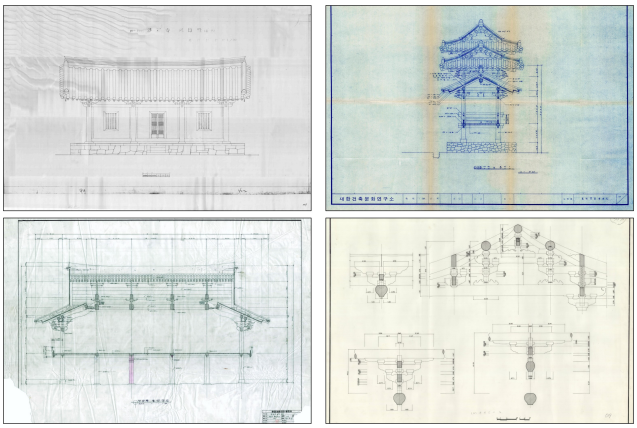


Figure 3. Examples of aged hand-drawn architectural drawings in the archive.

The degradation generator is learned in the second stage, at the same time while the encoder E_n is trained, but is detached in the latent space so that the backpropagation of G_n does not flow into E_n.

The generator G_n is utilized to produce an augmented version of a noisy drawing from a clean drawing even while the training is in progress. Because the intermediate activation of the generator G_c is fed to the generator, the generator starts to generate an authentic noisy drawing corresponding to the clean input drawing $\mathbf{x}_c$ after only a few steps in the training. We use clean drawing and either the corresponding synthesized or generated noisy drawings alternatingly to augment the training set used for the train of the latent space mapping by the encoder E_n.

3.4. Restoration of Noisy Drawings

Once the model is trained through the first and second stages, the restoration of noisy drawings is performed by encoding a noisy drawing with the encoder E_n and the vector quantization. These operations map the latent variables of the noisy drawing to those of a clean drawing and quantized using the codebook $\mathcal{C}$. Then the quantized latent variables are decoded with the generator G_c to restore a clean drawing.

4. Experiments

4.1. Drawings of Korean National Treasures

The archive of traditional wooden buildings drawings is available on the Cultural Heritage Administration website[1]. The dataset contains frontal elevations, side elevations, and floor plans of 55 wooden-built temples and Confucius academies (12 national heritages and 43 heritages) drawn by ten studios between the 1970s and 2000s. The

[1] http://english.cha.go.kr/cha/idx/SubIndex.do?mn=EN

resolutions of the drawings are between 3000×4000 and 8000×10000 pixels. Some drawings were digitized after they had aged in an inadequate environment and showed serious degradation. We collected 330 clean drawings and 350 aged drawings. These drawings are not paired, as archived drawings are either clean or aged. Examples of the aged drawings are shown in Fig. 3.

4.2. Network Architectures and Training

In the first stage, the VQ-VAE, i.e. the encoder E_c, generator G_c, and the codebook $\mathcal{C}$, is trained in an autoregressive setting using a set of clean drawings. We used 43000 patches of the size 256×256 cropped from 330 clean drawings for training. The codebooks were initialized randomly.

In the second stage, the latent space mapping by the encoder E_n is trained while the VQ-VAE, i.e., E_c, G_c, and $\mathcal{C}$, were frozen. The training requires a set of clean and noisy drawing pairs. For the clean drawings, the same 43000 clean patches from the 330 clean drawings were used. We used two types of noisy drawings: i) synthetic noisy drawings obtained by degrading clean patches using the method in [16, 19] with a background randomly chosen from 8000 noisy background patches, and ii) the outputs of the degradation generator. The training of the degradation generator requires a set of unpaired noisy drawings for the discriminator inputs. We used 47000 patches of the size 256 × 256 cropped from 350 aged drawings. The second stage of training starts with only the synthetic noisy drawings. After the degradation generator starts to generate authentic noisy drawings, the synthetic noisy drawings and the outputs of the degradation generator are alternately input. In our experiments, the outputs of the degradation generator were used after 20000 steps. Examples of synthetically degraded drawings are shown in Fig. 4. In the following section, examples of the degradation generator outputs are given with comparisons to other generative methods.

We considered two VQ-VAE configurations. The first configuration utilizes the hierarchical VQ [14] with one codebook for 64×64 latent variables and the other for 16×16 latent variables. The number of codewords is 4096 and 1024 for the two codebooks, respectively. The second configuration utilizes the residual VQ [11] with a codebook for 32×32 latent variables. The number of codewords is 8192. The number of latent variables are 256 for both configurations. The first and second configurations of the VQ-VAE consist of 70 and 74 million parameters. Both VQ-VAEs were trained using adam. The learning rate was set to $1e-3$. The training of the first and second stages were conducted with 120 and 60 epochs, respectively. All the methods in comparison were retrained with our dataset using the hyperparameters listed in the original papers but the epoch was set heuristically.

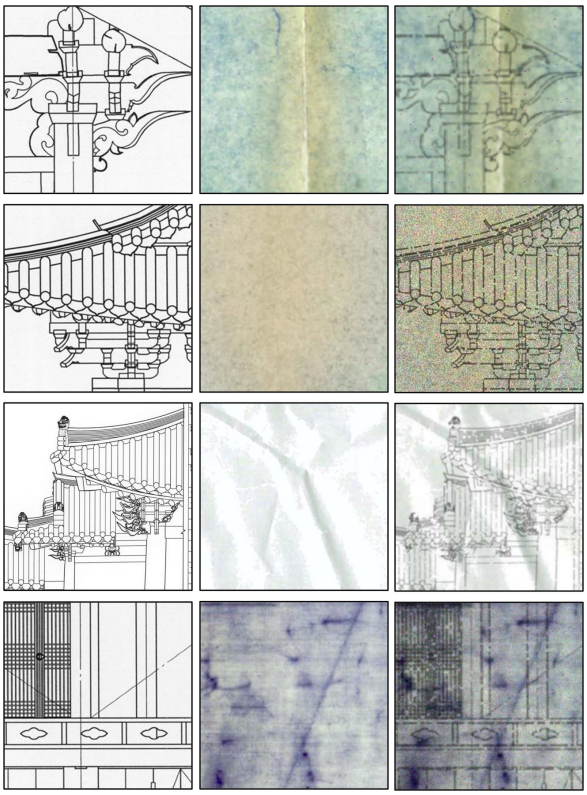


Figure 4. Example of synthetically degraded drawings. (a) clean drawings, (b) background images, and (c) degraded drawings.

Table 1. Quantitative comparison with existing methods for generating realistically degraded drawings.

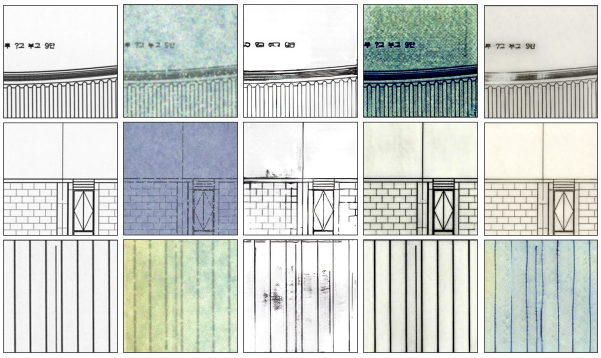
	KID↓	FID↓
CycleGAN [25]	0.035±0.001	59.979±0.08
MMA-CycleGAN [9]	0.033±0.001	54.862±0.15
Proposed (HVQ-DG)	0.015±0.001	29.663±0.05
Proposed (RVQ-DG)	0.023±0.001	41.159±0.06

In the following sections, the VQ configurations are denoted by HVQ and RVQ for the hierarchical and residual VQ, respectively, and the use of the degradation generator for producing drawing pairs in the training of the latent space mapping is denoted by DG.

4.3. Degradation Generator Performance

Table 1 shows the performance of the degradation generator. The performance of CycleGAN-based image-to-image translation methods [9, 25] is shown for comparison. All the methods were trained using clean drawing patches and aged drawing patches. FID and the kernel inception distance (KID) are used for the evaluation.

Figure 5 shows examples of synthetically degraded



(a) (b) (c) (d) (e)

Figure 5. Examples of drawings produced by various generation methods. (a) clean drawings, (b) synthetic degradation, (c) CycleGAN, (d) MMA-CycleGAN, and (e) proposed degradation generator.

drawings, drawings generated by the CycleGAN-based methods, and those by the degradation generator using clean drawings. The smeared, broken, and disposition lines were more authentically reproduced by the degradation generators.

4.4. Restoration Performance

The restoration performance of the proposed method is reported in Tab. 2. The VQ configurations are denoted by HVQ and RVQ for the hierarchical and residual VQ, respectively, and the use of the degradation generator in the training is denoted by DG. For comparison, we reported the performance of other learning-based restoration methods: restoration network with line detection sub-network [16], old photo restoration network with deep latent space translation [19], dual adversarial networks [20], and pixel-wise dilation filtering [21]. For supervised learning, synthetic noisy drawings [16, 21] were used to pair corresponding clean drawings. The peak signal-to-noise ratio (PSNR), the structural similarity measure (SSIM), the learned perceptual image patch similarity (LPIPS) [24], and the Frechet inception distance (FID) [7] are used for the evaluation. The up and down arrows indicate higher or low the better for the measures. The proposed methods provided significant improvement in all the measures.

Examples of restored synthesis drawings are shown in Fig. 7. The synthetically degraded drawings and the clean drawings are also shown. There are some details lost and noise left in the results by the compared methods. In comparison, the proposed methods showed improvement. However, all the methods showed relatively good performance with the synthetically degraded drawings, because they are trained to restore such synthetic degradations.

When the restoration methods were applied to restore archived aged drawings, there were significant differences

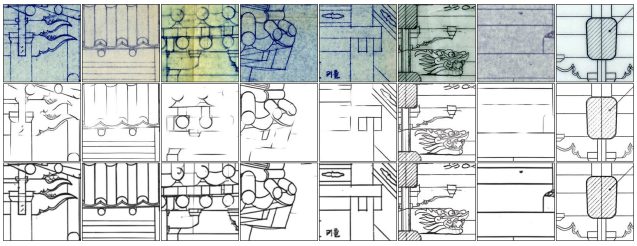


Figure 6. Ablation study with/without augmentation by degradation generator on restoration performance. From top to bottom: aged archived drawings, without DG, and with DG.

Table 2. Quantitative comparison of restoration performance with existing methods.

	SSIM↑	PSNR↑	LPIPS↓	FID↓
SASAKI *et al.* [16]	0.8870	21.32	0.1977	98.677
WAN *et al.* [19]	0.9548	22.73	0.0632	36.090
YUE *et al.* [20]	0.9534	23.83	0.0713	42.121
Guo *et al.* [21]	0.9616	24.78	0.0585	37.025
Proposed (HVQ)	0.9683	25.89	0.0559	36.812
Proposed (RVQ)	0.9613	24.96	0.0531	36.468
Proposed (HVQ-DG)	0.9673	25.82	0.0564	36.590
Proposed (RVQ-DG)	0.9576	24.51	0.0542	33.500

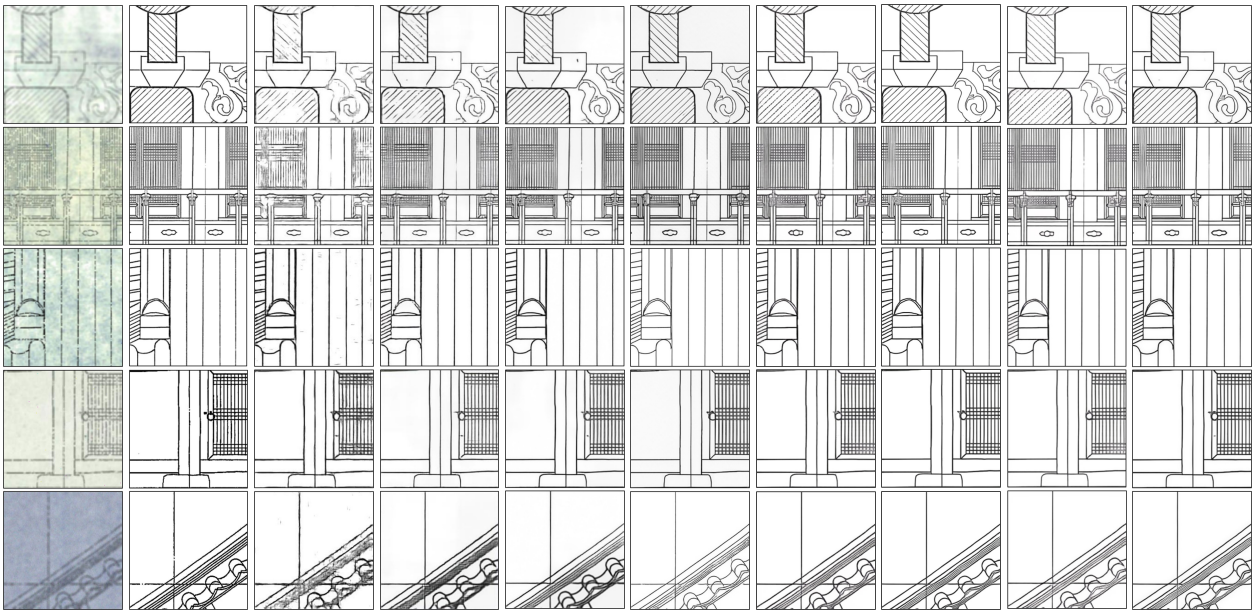
in the restoration performance. Figure 8 shows examples of restored aged drawings. The smeared ink, fading fine and packed lines, or even the ink smudged from the adjacent drawings causes difficulties for the other methods. They fail to clean up the smearing, reconnect the broken lines, or suppress the noise in the background. The drawing restored by the proposed methods showed exceptional quality.

The significant performance gain by the proposed method is mainly due to the augmentation using the degeneration generator. Figure 6 shows the ablation study on the restoration performance when the latent space mapping is trained with and without the degradation generator. Without the degradation generator, the proposed method misses fading and packed lines and fails to restore them. With the data augmentation by the degradation generator, the latent space mapping is properly trained for more accurate restoration in subjective evaluations.

Figure 9 shows examples of the entire restored drawings with zooming on two positions for full inspection. Results by the methods in [19, 20] and the proposed methods are shown.

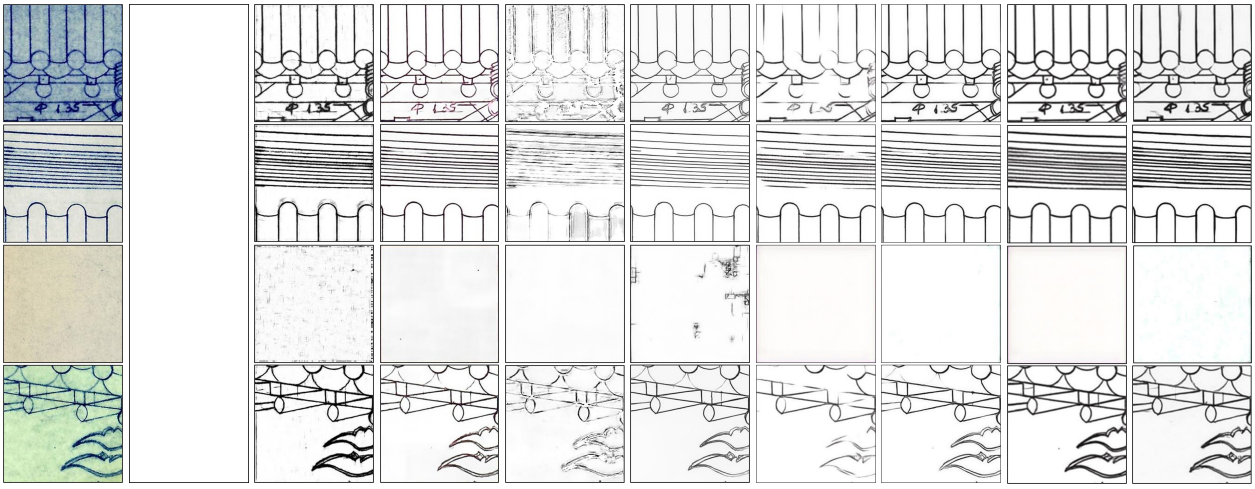
5. Conclusion

Aged hand-drawn drawings of architectural heritage are restored using a novel latent space mapping methods. The accurate representation of clean drawings was learned with a set of clean drawings. A generator that generates authentic noisy drawings from the residual latent space is trained us-



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 7. Restoration results of synthetically degraded drawings. (a) synthetically degraded images, (b) ground truth, (c) Sasaki *et al*., (d) Yue *et al*., (e) Guo *et al*., (f) Wan *et al*., (g) Proposed (HVQ), (h) Proposed (RVQ), (i) Proposed (HVQ-DG), and (j) Proposed (RVQ-DG).



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 8. Restoration results of aged drawings. (a) aged archived drawings, (b) no ground truth available, (c) Sasaki *et al*., (d) Yue *et al*., (e) Guo *et al*., (f) Wan *et al*., (g) Proposed (HVQ), (h) Proposed (RVQ), (i) Proposed (HVQ-DG), and (j) Proposed (RVQ-DG). For aged archived drawings, the proposed method with DG showed better restoration performance than the other methods in comparison.

ing a set of noisy drawings. An accurate latent space mapping from noisy to clean drawings was learned using the degradation generator for data augmentation. The proposed method performed exceptional restoration, generating clean drawings from the codebook, and reported a significant gain in quantitative measures and qualitative inspections over other existing learning based methods.

Acknowledgements

This research was supported by 2023 Cultural Heritage Smart Preservation & Utilization R&D Program of Cultural Heritage Administration, National Research Institute of Cultural Heritage (Project Name: Development of AI based CAD conversion technology for traditional architecture drawing images, Project Number: 2023A02P03-001, Contribution Rate: 100%)

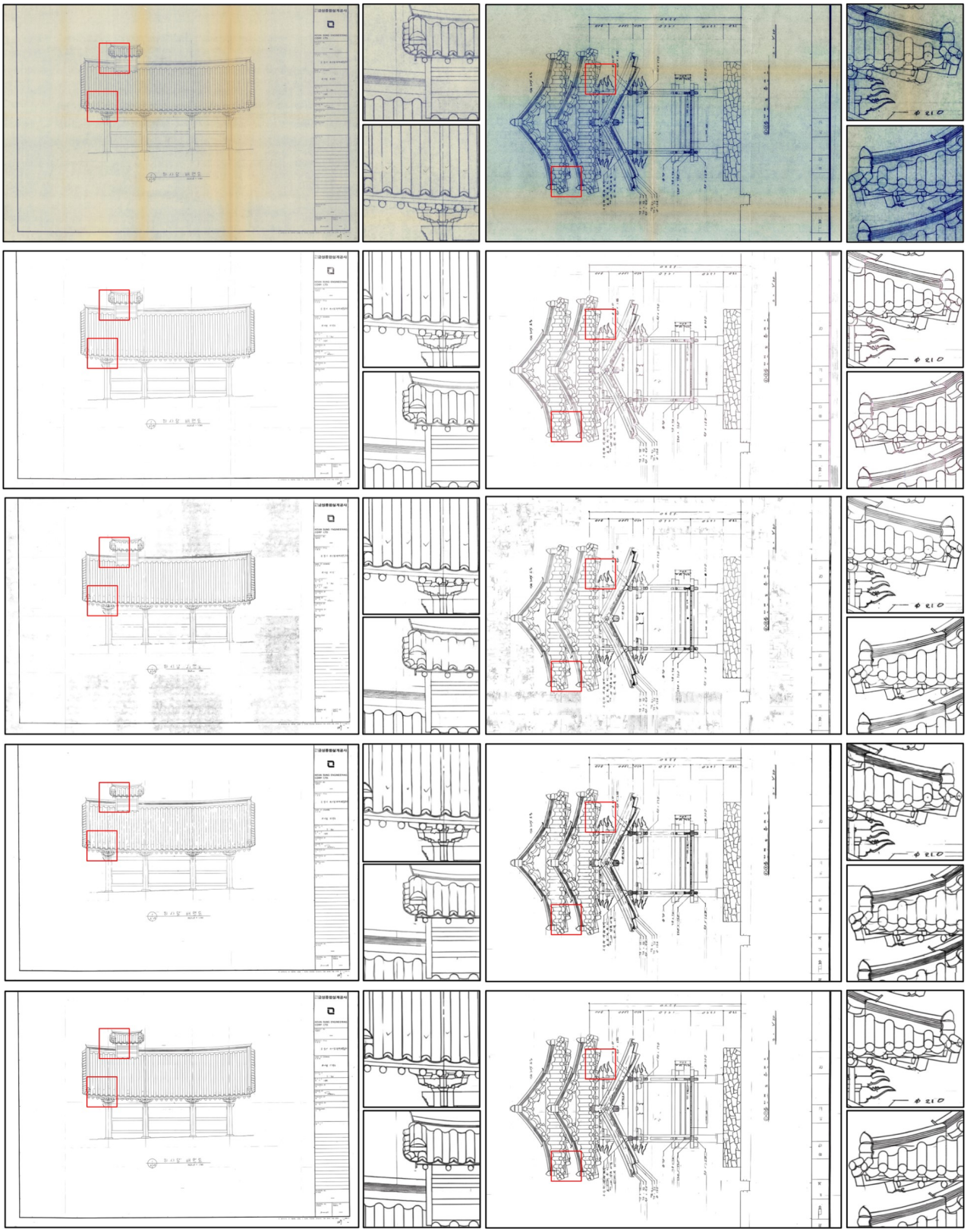


Figure 9. Restoration results of archived aged drawings with zooming two positions. From top to bottom: archived drawings, Yue *et al*., Wan *et al*., proposed (HVQ-DG), and proposed (RVQ-DG).

References

[1] Nakkwan Choi, Seungjae Lee, Yongsik Lee, and Seungjoon Yang. Denoising traditional architectural drawings with image generation and supervised learning. *Journal of architectural history*, 31(1):41–50, 2022. 2

[2] Umberto Cugini, G Ferri, Piero Mussio, and Marco Protti. Pattern-directed restoration and vectorization of digitized engineering drawings. *Computers & graphics*, 8(4):337–350, 1984. 2

[3] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *ICCV*, pages 764–773, 2017. 4

[4] Vage Egiazarian, Oleg Voynov, Alexey Artemov, Denis Volkhonskiy, Aleksandr Safin, Maria Taktasheva, Denis Zorin, and Evgeny Burnaev. Deep vectorization of technical drawings. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 582–598. Springer, 2020. 2

[5] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *CVPR*, pages 12873–12883, 2021. 2

[6] Jean-Dominique Favreau, Florent Lafarge, and Adrien Bousseau. Fidelity vs. simplicity: a global approach to line drawing vectorization. *ACM Transactions on Graphics (TOG)*, 35(4):1–10, 2016. 2

[7] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, volume 30, 2017. 6

[8] Rik D.T. Janssen and Albert M. Vossepoel. Adaptive vectorization of line drawing images. *Computer vision and image understanding*, 65(1):38–56, 1997. 2

[9] Wei Ji, Jing Guo, and Yun Li. Multi-head mutual-attention cyclegan for unpaired image-to-image translation. *IET Image Processing*, 14(11):2395–2402, 2020. 5

[10] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, pages 8346–8355, 2020. 1

[11] Doyup Lee, Chiheon Kim, Saehoon Kim, Minsu Cho, and Wook-Shin Han. Autoregressive image generation using residual quantization. In *CVPR*, pages 11523–11532, 2022. 5

[12] Chengze Li, Xueting Liu, and Tien-Tsin Wong. Deep extraction of manga structural lines. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017. 2

[13] Van Den Oord, Aaron, and Oriol Vinyals. Neural discrete representation learning. In *NeurIPS*, volume 30, 2017. 1, 2

[14] Ali Razavi, Aaron van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae-2. In *NeurIPS*, volume 32, 2019. 2, 5

[15] Kazuma Sasaki, Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Joint gap detection and inpainting of line drawings. In *CVPR*, pages 5725–5733, 2017.

[16] Kazuma Sasaki, Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Learning to restore deteriorated line drawing. *The Visual Computer*, 34(6):1077–1085, 2018. 1, 2, 5, 6

[17] Edgar Simo-Serra, Satoshi Iizuka, and Hiroshi Ishikawa. Mastering sketching: adversarial augmentation for structured prediction. *ACM Transactions on Graphics (TOG)*, 37(1):1–13, 2018. 2

[18] Edgar Simo-Serra, Satoshi Iizuka, Kazuma Sasaki, and Hiroshi Ishikawa. Learning to simplify: fully convolutional networks for rough sketch cleanup. *ACM Transactions on Graphics (TOG)*, 35(4):1–11, 2016. 2

[19] Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Fang Wen, and Jing Liao. Old photo restoration via deep latent space translation. *IEEE TPAMI*, pages 1–1, 2022. 2, 5, 6

[20] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *ECCV*, pages 41–58, Cham, 2020. Springer. 2, 6

[21] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *AAAI*, pages 1487–1495, 2021. 1, 6

[22] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 1, 2

[23] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 26(7):3142–3155, 2017. 1

[24] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 6

[25] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pages 2223–2232, 2017. 5