Check for updates

Received: 28 February 2025 | Revised: 1 May 2025 | Accepted: 13 June 2025

DOI: 10.4218/etrij.2025-0113

ORIGINAL ARTICLE

ETRI Journal WILEY

RICH: Coin data restoration using inpainting for cultural heritage analysis

Seohyun Baek[1] | **Jae-Ho Lee**[1] | **Sangdoo Ahn**[2,3] | **Jongwon Choi**[2,4]

[1]Content Research Division, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

[2]Department of Cultural Properties, Chung-Ang University, Seoul, Republic of Korea

[3]Department of Chemistry, Chung-Ang University, Seoul, Republic of Korea

[4]Department of Advanced Imaging, The Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul, Republic of Korea

Correspondence
Jongwon Choi, Department of Advanced Imaging, The Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul, Republic of Korea.
Email: choijw@cau.ac.kr

Funding information
Ministry of Science and ICT (MSIT, Republic of Korea) & Gwangju Metropolitan City; Ministry of Culture, Sports and Tourism and Korea Creative Content Agency, Grant/Award Number: RS-2023-00219579; Institute of Information & Communications Technology Planning & Evaluation (IITP), Grant/Award Number: RS-2021-II211341

Abstract

The rapid development of artificial intelligence technology has expanded its use in the restoration of cultural heritage. Although research on the digital restoration of ancient coins is well established, most studies have focused on Western currencies—such as coins from the Roman Empire and Greece—with little prior research being conducted on ancient Korean coins. This study uses Sangpyeongtongbo coins—which, among ancient currencies, were widely circulated throughout Korea—as its subject. For this research, a new dataset is constructed using Sangpyeongtongbo specimens from museums and research institutes. The Restoration using Inpainting for Cultural Heritage (RICH) method is developed based on stable diffusion and attempted digital restoration, to successfully recover damaged Chinese inscriptions on the surface of coins. To validate the restored image, image analysis of the undamaged and restored inscriptions is conducted, with the results demonstrating similarity rates of 97.10% for entire inscriptions and 83.81% for restored areas. These results demonstrate the feasibility of using artificial intelligence for digital restoration.

KEYWORDS
cultural heritage, diffusion, digital restoration, inpainting, sangpyeongtongbo

1 | INTRODUCTION

Cultural heritage, as a product of history and tradition, is essential to understand the unique cultural characteristics of a nation and the evolution of its people's lives [1]. Cultural heritage holds enormous historical value as it reflects the overall social, cultural, and political contexts of an era while simultaneously encompassing artistic, symbolic, and socioeconomic values [2]. In the context of cultural heritage, ancient coins serve not only as a currency but also as crucial historical materials that reflect the economics, politics, society, and technology of their time. Ancient coins exhibit various changes in material composition, design elements, metal content, and symbolic expression [3]. However, ancient coins can be damaged by environmental factors, careless management, and physical or chemical factors, necessitating preventive conservation and restoration work.

This is an Open Access article distributed under the term of Korea Open Government License (KOGL) Type 4: Source Indication + Commercial Use Prohibition + Change Prohibition (http://www.kogl.or.kr/info/licenseTypeEn.do).
1225-6463/$ © 2025 ETRI

Metallic cultural heritage items such as coins are particularly vulnerable to physical damage and wear, being prone to oxidation and corrosion when exposed to environments such as soil or the seabed for extended periods. Such damage can cause a loss of surface information, making it difficult to identify and classify ancient coins, and ultimately limiting our understanding of their historical value and characteristics. Consequently, research and development for the preservation and restoration of ancient coins is essential.

The preservation and restoration of ancient coins require a systematic scientific approach to analyze the remaining information on their surfaces and restore lost elements. For this purpose, traditional methods—such as X-rays, computed tomography (CT), X-ray fluorescence (XRF), and energy dispersive spectroscopy (EDS)—have been employed to identify their internal structures and compositional elements [4–6]. However, with the rapid advance of digital technology, restoration techniques using artificial intelligence (AI) have garnered increasing attention. These technologies have influenced the field of cultural heritage restoration and management research, with the relationship between cultural heritage and digital technology continuing to evolve [7]. For example, there have been cases that have applied various technologies, including 3D object reconstruction, virtual reality, mathematical algorithms (incorporating diffusion models), and other digital restoration techniques [8–10].

Recently, generative AI—which can create new 2D or 3D images using algorithms such as the generative adversarial network (GAN) [11]—has garnered considerable attention [12]. GAN-based methods have demonstrated the potential to improve the appearance and quality of damaged artifacts, restoring degraded objects to physically enhanced forms [12]. Moreover, research has been conducted on developing algorithms that use adversarial learning centered on a GAN to learn the degradation process of ancient coins and restore their pre-damaged appearance [13].

Research on digital coin restoration continues to advance, but studies of ancient coins in Asia (including Korea) remain relatively scarce. Although studies on ancient numismatics using computer vision and machine learning techniques have demonstrated the feasibility of digital restoration technologies, they have focused predominantly on datasets of ancient Roman coins [13]. These coins were produced in various forms, distinguished by the depiction of figures, animals, or other symbols, whereas Asian coins were categorized primarily based on Chinese inscriptions. This distinction poses additional challenges in identifying the coin types when their surfaces are damaged. In this context, this study selected the Sangpyeongtongbo, a representative ancient Korean coin, as the primary research subject.

In this study, we photographed images of Sangpyeongtongbo coins to create a novel dataset, based on which we proposed an algorithm to predict and digitally restore damaged Sangpyeongtongbo coins. The proposed method can digitally restore lost information without causing direct physical damage to the Sangpyeongtongbo coins. Through this, we aim to support researchers in more accurately analyzing the surface details of damaged Sangpyeongtongbo coins, which can be difficult to identify, and contribute to improving preservation efficiency by using the restored information as reference material for conservation work. Moreover, the proposed algorithm is expected to demonstrate the scalability and potential application of digital restoration technology, as it can be applied to Sangpyeongtongbo coins as well as to the restoration research of other types of ancient coins. The code is available on GitHub.[1]

2 | RELATED WORK

Recent advances in image restoration and generation techniques have introduced various deep-learning-based methods that aim to recover both the structural consistency and fine textures of damage-related images. Transformer-based inpainting methods [14], multi-stage structure–texture integration models [15], super-resolution using pre-trained diffusion models [16], and CNN-based dehazing networks [17, 18] have been developed as alternatives or enhancements to GANs. Despite these diverse methods, GANs remain central to image generation and inpainting models. This section reviews key GAN-based models and deep-learning techniques for coin restoration.

2.1 | Generative models

Generative models are neural network models with the ability to create new content that has not existed previously, thereby producing realistic new images within the domain of their training dataset [19]. Therefore, it is necessary to develop metrics that measure their closeness to existing models and calculate gradients to minimize these differences [20]. Notable generative models include variational autoencoders (VAEs) [21], which learn latent representations of data and generate new data based on these representations, and GANs [11], which generate high-quality data through competition between two

[1]GitHub repository: https://github.com/Seohyun-Baek/RICH

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

neural networks—that is, a generator and a discriminator. Diffusion models [22] have also become an active area of research.

The VAE can learn the characteristics of training data using latent variables and generate data similar to the training data [21]. It can calculate the mean vector ($\boldsymbol{\mu}$) and variance vector ($\boldsymbol{\sigma}$) from the input using an encoder, before sampling latent variables according to probability and reconstructing the original data using a decoder. One notable characteristic is that it can generate continuously changing data by adjusting the latent variables [23]. This results in different outputs for the same input, demonstrating its ability to extract core features with minimal noise. Additionally, it is guaranteed to operate even with unknown inputs. Consequently, VAEs can be used in various applications, proving to be particularly effective in image generation, data compression, and feature extraction applications.

The GAN, proposed in 2014, is a novel type of neural network architecture [11] that can learn image characteristics and generate highly realistic images, videos, music, and text based on the learned patterns [19]. It is a generative model in which two neural networks—that is, a generator and discriminator—learn through competition with each other and is frequently used in image generation applications [23]. Initially, a multilayer perceptron was used to create the generator and discriminator; later, a deep convolutional GAN was proposed, which enhanced the model's predictive capabilities by using convolutional layers in the discriminator network. GANs are widely used in high-resolution image generation, image adaptation (through style imitation), and image synthesis applications. Additionally, various GAN-based technologies have been developed, including text-to-image generation and automatic design technologies. Zhang and others [24] introduced the MMInvertFill model, a multimodal GAN inversion method that integrated RGB, edge, and semantic segmentation inputs to enhance both structural consistency and fine-detail recovery; Lu and others [25] proposed the EXE-GAN model, which was an exemplar-guided facial inpainting model that achieved realistic restoration using style similarity and spatial adaptation techniques; and Ning and others [15] presented a two-stage GAN-inspired model for cultural heritage restoration that separated structural and texture reconstructions to achieve high-fidelity results.

The diffusion model is a deep generative model for data generation and is currently one of the best-performing models in the image-generation field [22]. It is based on two main processes—that is, the forward process (diffusion process), which gradually adds noise to real data until they become completely noisy, and the reverse process, which gradually restores the data from the noise. By learning these processes, the desired images, texts, and graphs, can be generated from random noise. For example, the inpainting diffusion model is based on a GAN and has been used to generate and modify images based on the input text [22]. This model converts text into images and uses masks to maintain structural consistency and texture clarity during the inpainting process. Methods such as RePaint have used image inpainting methods based on denoising diffusion probabilistic models (DDPMs). These methods generate various output images without modifying the DDPM network. Overall, diffusion models have demonstrated superior performance compared with GAN-based methods in image generation applications.

2.2 | Coin restoration using AI

Most research conducted to restore coins has been conducted using traditional restoration methods [26–29]. Dumitriu and others [27] selected nine ancient Roman coins with varying degrees of corrosion, and production years to present a preliminary analysis of the stages related to the preservation and restoration of metal artifacts and determined the authenticity of the coins through information such as their morphological characteristics, weight, and dimensions. However, ancient coins that are imaged today were produced thousands of years ago, and even if each coin undergoes the same casting process, it is impossible to know its exact appearance before being damaged because of the primitive casting processes used at the time [13]. Therefore, it can be difficult for people without professional knowledge to accurately identify heavily corroded ancient coins. Consequently, this study aims to introduce and evaluate cases in which deep learning based on 2D images can be applied to ancient-coin restoration.

Zachariou and others [13] aimed to synthesize images similar to their original state based on damaged ancient Roman coin images using deep-learning-based algorithms. Specifically, they developed a cycle-consistent generative adversarial network (CycleGAN) that could phenomenologically model the process of damage to ancient coins using adversarial learning centered on GANs. Inspired by the PatchGAN model [30], the CycleGAN comprises five layers using two ResNets and employs two generators and two discriminators. Each generator learns to map from one domain to another. The first generator learns how to generate synthetic images of extremely fine coins using fine coin input images, whereas the second generator operates in the opposite direction, learning to model the degradation

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

process that ancient coins experience over time. Through the trained algorithm, various missing details of symbols minted on the coin surfaces could be successfully reconstructed, confirming that the algorithm correctly learned the range of appearances present on the coin surfaces and appropriately applied them in the visual context. However, although this digital restoration technique primarily enhanced the surface details of corroded coins, such as facial features, it demonstrated no improvement in other forms of damage, such as missing structural components, cracks, or discoloration [12].

Altaweel and others [12] built a dataset by collecting ancient Roman coin images from Nomisma-based websites and attempted to reconstruct damaged areas (including cracks), using the CycleGAN model. They confirmed that the ability to improve damaged coins was enhanced by generating synthetic images that were similar to those of real coins. However, the results varied depending on the degree of damage to the coin surface. When damage to the coin surface was less than 30%–40%, an overall improvement was evident, but when more than 40% of the surface information was damaged, it proved to be difficult to reconstruct the coin surface. Moreover, they found that changes in the color and lighting conditions affected the model's learning speed and accuracy.

3 | METHODOLOGY

3.1 | Dataset acquisition

Sangpyeongtongbo, whose name means "currency that maintains a constant value in circulation," was first minted in 1678 (the fourth year of King Sukjong's reign) and remained in circulation throughout the nation for approximately 200 years until 1888 (the 25th year of King Gojong's reign), serving as the representative currency of the Joseon Dynasty [31]. The surfaces of these coins contained crucial information regarding their minting period, offices, and mints, making them valuable heritage artifacts for archaeological and cultural research [32].

The front of each coin was inscribed with "常平通寶 (Sangpyeongtongbo)," whereas the reverse side featured various identifiers, including abbreviations of the manufacturing office at the top, Thousand Character Classic characters indicating the mint location at the bottom, and numbers and symbols on the sides. The minting system was extensive, comprising six central institutions, including Ho-Jo (戶), Hunryun-Dogam (訓), Geumwiyeong (禁), and Yeoyong-Chong (營), as well as provincial institutions such as Gaeseong-Gamyong (開), Pyongan-Gamyong (平), Jeolla-Gamyong (全), and Gyeongsang-Gamyong (尙) [4]. Each minting office managed approximately 20 mines, resulting in approximately 4000 different varieties of Sangpyeongtongbo coins through various combinations of mints, inscriptions, calligraphic styles, symbols, and ordering systems [33].

In this study, we constructed a comprehensive dataset by examining Sangpyeongtongbo coins from multiple sources, including the Gongju National Museum, the National Research Institute of Maritime Heritage, and the Seoul National University Museum, along with privately owned specimens. A total of 434 Sangpyeongtongbo specimens were photographed using a Nikon D850 DSLR camera equipped with an AF-S Micro Nikon 105 mm 1:2.8G ED lens. Their distribution across different institutions is summarized in Table 1 and categorized according to their minting offices, as shown in Table 2.

The dataset was split into training and validation datasets, with 70% of the data being used for training and 30% being used for validation, maintaining consistent class-distribution ratios across both datasets. To validate the effectiveness of the proposed algorithm, we used undamaged coins from the dataset. Using the MI-GAN

TABLE 1 Distribution of the Sangpyeongtongbo dataset by institution.

Repository	Gongju National Museum	National Research Institute of Maritime Heritage	Seoul National University Museum	Private ownership	Total
Amount	24	47	50	50	434

TABLE 2 Distribution of the Sangpyeongtongbo dataset by minting authority.

Pyongan-Gamyong				Ho-Jo	Hunryun-Dogam
Chinese inscriptions	Left sun	Right sun	Number inscription		
50	50	50	50	50	50

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

model [34], the first generative image inpainting network known to produce high-quality results with minimal computational overhead, we removed and restored various inscribed characters. The generated images served as training data for our diffusion networks, enabling the restoration of damaged characters in the input images.

3.2 | Generation model for restoration

The main contribution of this research is the confirmed generation of Chinese inscriptions on Sangpyeongtongbo surfaces using AI models and the demonstrated feasibility of digital restoration. We explored the optimal base architecture for this purpose and developed a specialized algorithm for Sangpyeongtongbo coins. The algorithm developed was named Restoration using Inpainting for Cultural Heritage (RICH), signifying its role in restoring images of ancient coins.

During the model selection process, a comparative analysis of the GAN and diffusion models was conducted [35]. Although the GAN offers advantages, such as high-quality generated images and rapid inference after training completion, it has the limitation of requiring separate training for each new dataset [22]. By contrast, diffusion models can be easily generalized, as they can use pre-trained models for inpainting without modification [22] and demonstrate excellent performance, even in limited data environments. After considering these characteristics, we adopted Stable Diffusion [36] as the base model because it could be effectively applied, even with a limited Sangpyeongtongbo dataset (Figure 1).

The RICH model adopted U-Net [37] as its backbone for denoising and improved its inference speed through single-step denoising. The U-Net model is structured using convolutional layers in a 'U' shape and features identical input and output image dimensions, which were used in the RICH model to maintain the same input and output shapes. For efficient model training, the images were compressed using a VAE before being input to the U-Net model. Moreover, to minimize information loss during dimension reduction, they were restored to their original form using a VAE decoder [21]. By applying low-rank adaptation (LoRA) adapters [38] to the U-Net and VAE to enable the training of specific layers, rapid and memory-efficient training could be achieved. Training was conducted using a batch size of four and a weight decay of 0.01 to prevent overfitting. To ensure stability during the initial training phase, the warm-up steps were set to 500 and the gradients were updated for each batch over 100000 training steps.

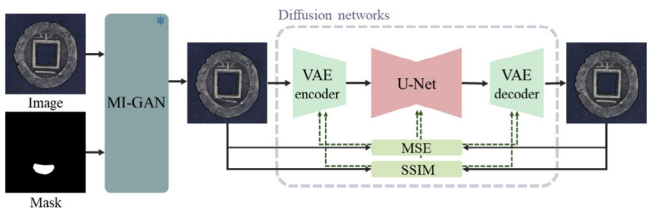


FIGURE 1 Overall RICH architecture: RICH is a diffusion-based inpainting model that uses a VAE and U-Net. The VAE identifies the features of the training data using latent variables and generates data similar to the training data. U-Net, built with stacked convolutional layers, helps preserve detailed image information and prevents loss during the reconstruction process. RICH, Restoration using Inpainting for Cultural Heritage; VAE, variational autoencoders.

Furthermore, the mean squared error (MSE) loss [39] and structural similarity index measure (SSIM) loss [40] were combined to optimize the image quality and structural similarity simultaneously. Although the MSE loss could generate visually unnatural images owing to its focus on minimizing pixel-level differences, the SSIM loss could produce more natural and high-quality images by emphasizing the structural patterns and textures. By combining these complementary characteristics, we could maintain pixel-level accuracy through MSE loss, while achieving better structural similarity through SSIM loss.

The MSE loss can be calculated by squaring the difference between the predicted and actual values, computing their mean and assigning equal weights to all data points to prevent errors from being biased in a particular direction. This can be effective in training the model to ensure that the predicted values do not deviate significantly from the actual values and is used to minimize the pixel-level differences during the image generation or restoration processes.

$$\text{MSE loss} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{1}$$

where y_i denotes the actual pixel intensity, $\hat{y}_i$ denotes the model's predicted pixel intensity, and n denotes the total number of data samples.

The SSIM loss is a loss function designed to reflect the human visual system and can be used as a loss function for measuring the similarity between two images or patches when calculating the disparity in image restoration or depth estimation processes. Its purpose is to minimize the difference between the generated and target images during model training and to guide the model to generate structurally similar images during learning processes (such as image generation, restoration, and

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

transformation). Generally, one SSIM value is converted into a loss value that must be minimized.

$$\text{SSIM loss} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (2)$$

where x and y denote the corresponding windows of the two images being compared, μ_x and μ_y denote their respective mean values, σ_x and σ_y denote the standard deviations, σ_{xy} denotes covariance, and C_1 and C_2 denote constants used to prevent the denominator from becoming zero.

3.3 | Evaluation metric

Quantitative evaluation was conducted from two perspectives. First, we measured the SSIM of the entire image to evaluate whether the restored image maintained the characteristics of the original image. Second, we separately measured the SSIM specifically for the damaged and restored areas to analyze the restoration quality in detail.

In this study, we used the SSIM to objectively evaluate and quantify the quality of restored images [40]. The SSIM is an evaluation metric that reflects how the human visual system recognizes structural information in images by comprehensively comparing the luminance, contrast, and structure of two images. It takes values between −1 and 1, where values closer to 1 indicate a greater structural similarity between the original and restored images.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (3)$$

$$\text{SSIM}(\overline{x}, \overline{y}) = \frac{(2\mu_{\overline{x}}\mu_{\overline{y}} + C_1)(2\sigma_{\overline{xy}} + C_2)}{(\mu_{\overline{x}}^2 + \mu_{\overline{y}}^2 + C_1)(\sigma_{\overline{x}}^2 + \sigma_{\overline{y}}^2 + C_2)}. \quad (4)$$

To ensure a robust evaluation of the restored regions, Equations (3) and (4) define the variable x as the representation of the overall image, whereas $\overline{x}$ specifically highlights the restored regions. This distinction enables a clear comparison between the complete image and its reconstructed portions. The SSIM calculation follows this distinction by incorporating localized statistical measures, ensuring that the structural integrity of the restored regions is properly assessed.

In this process, the mean (μ) and variance (σ) of both x and $\overline{x}$ are computed using a Gaussian window to capture the localized luminance and contrast variations. We employed a Gaussian filter with a window size of 11×11 and a standard deviation of 1.5 to achieve optimal local feature extraction while maintaining computational efficiency.

The SSIM formulation includes two stability constants (C_1 and C_2) which prevent division by zero. These constants are defined as $C_1 = (K_1 \cdot L)2$ and $C_2 = (K_2 \cdot L)2$, where L denotes the pixel intensity range, and K_1 and K_2 denote empirically determined coefficients. The standard values for these parameters are $K_1 = 0.01$ and $K_2 = 0.03$, ensuring numerical stability while maintaining sensitivity to structural differences.

To enhance visual understanding, illustrative examples of these restored regions, along with their SSIM evaluations, are shown in Figure 2.

4 | RESULTS AND DISCUSSION

4.1 | Implementation details

Stable Diffusion ver. 2.1 was used as the base generator. The input images were resized to 512×512 crops in four batches using the H100 GPU and optimized using AdamW with a learning rate of 0.0001, based on a pretrained model. The training dataset comprised 508 Sangpyeongtongbo images. The RICH model was trained for 60 000 epochs in approximately 15 h.

The architectural details were as follows: the VAE used in this study first transformed an input RGB image at a resolution of 512×512 into 128 channels through an initial convolutional operation and then sequentially expanded it to 256 and 512 channels through three downsampling blocks before finally compressing it into a latent space with four channels and a spatial resolution of 64×64. The decoder received this latent

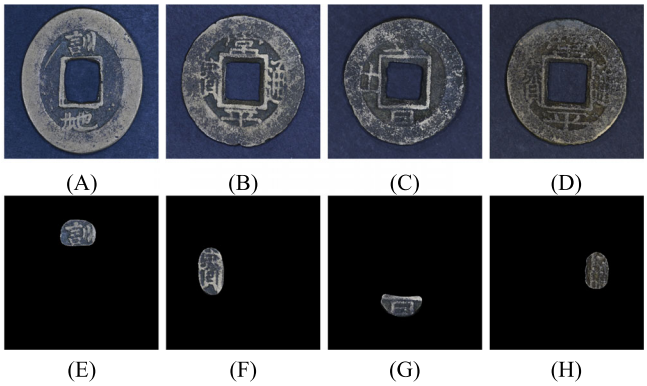


FIGURE 2 Example images used for the SSIM evaluation. (A–D) The entire actual image x, (E–H) Images show x, which focuses on local areas to be restored.

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

representation and restored it to a 3-channel RGB image through the upsampling stages of the 512 256 and 128 channels, whereas intermediate feature maps from the encoder were processed through dedicated skip convolution layers and integrated into the decoder. The U-Net block took a four 4-channel 64 × 64 latent representation as input, expanded to 320 channels through an initial convolutional operation, and reconstructed the latent representation through four downsampling stages, a bottleneck block, and four upsampling stages.

Each stage included residual blocks and cross-attention modules conditioned on text features, and the entire network was based on the pre-trained Stable Diffusion Turbo model, where all convolutional and residual blocks used the SiLU activation function by default, with memory optimization achieved through xFormers-based cross-attention and efficient fine-tuning enabled by the LoRA method.

4.2 | Digital restoration results

The experimental results demonstrated that the RICH model could effectively restore Chinese inscriptions on damaged ancient coins. Figure 3 shows a close-up comparison between the damaged and restored regions,



FIGURE 3 Close-up view of the restored image highlighting detailed inscription reconstruction.

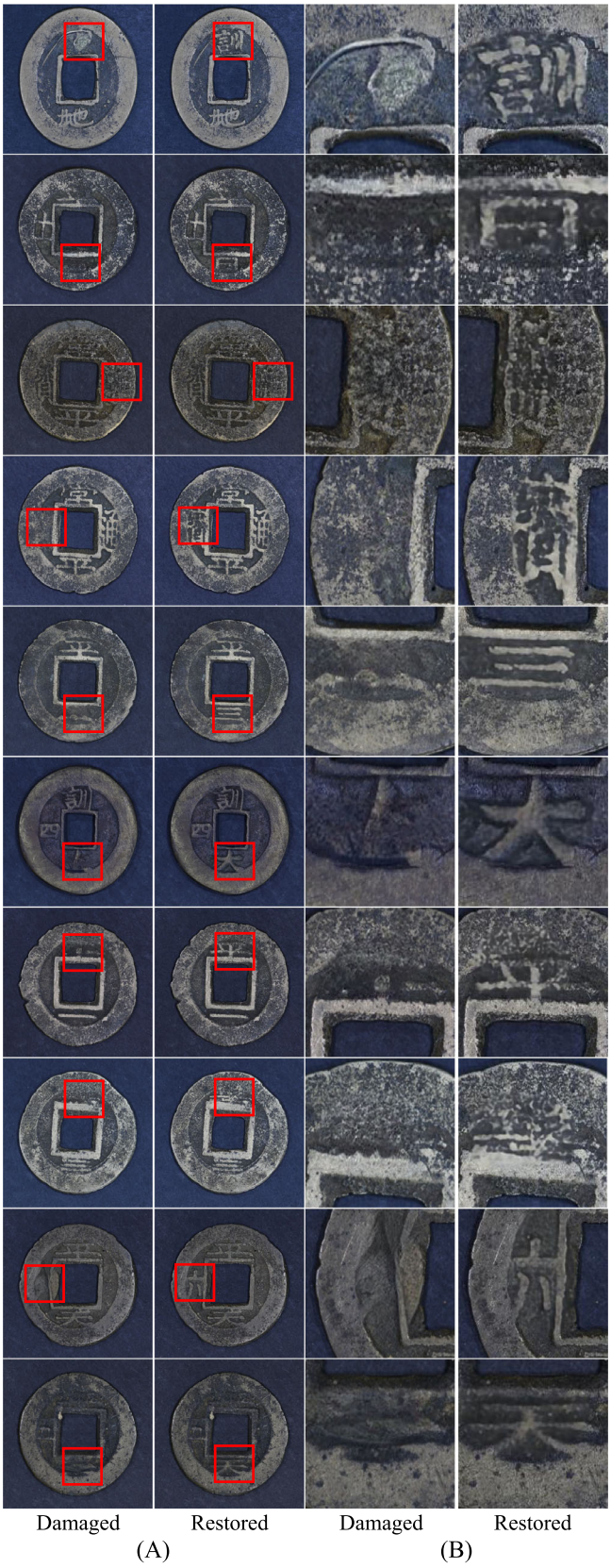


FIGURE 4 Sangpyeongtongbo restoration results using the RICH model. (A) Full image of Sangpyeongtongbo, (B) Detailed image of the restored part from image (Figure 4A).

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

clearly showing the successful reconstruction of the fine-grained inscription strokes.

As illustrated in Figure 4, the model enhanced the corroded and worn surfaces of the Sangpyeongtongbo specimens and achieved accurate restoration of the complex structures and visual characteristics of the original inscriptions. The restored images not only recovered the deteriorated sections of the coins but also improved their overall readability and preserved their aesthetic integrity, demonstrating the practical applicability of the proposed method. Although slight blurriness remained in certain restored areas, we expect the performance to improve with additional training data in future experiments.

Quantitative evaluation also supported these findings, with the overall SSIM of the restored images recorded at 97%, indicating a high structural similarity to the original images. Moreover, the SSIM for the damaged regions was 84%, clearly demonstrating successful restoration of the degraded areas.

4.3 | Ablation studies

We conducted an ablation study to evaluate the effectiveness of the different loss—function combinations in the RICH model. As shown in Table 3, each configuration was evaluated using the MSE [41–43], peak signal-to-noise ratio (PSNR) [41–43], learned perceptual image patch similarity (LPIPS) [41–43], and MS-SSIM metrics.

The combination of the MSE and SSIM yielded the highest PSNR value (29.88) and highest MS-SSIM score (0.9761), indicating robust performance in both pixel-level accuracy and structural consistency. By contrast, the mean absolute error (MAE) and LPIPS combination achieved the lowest LPIPS score (0.0997), suggesting superior perceptual quality.

The combination of the MAE and SSIM performed the worst across all metrics and was used as the baseline. These results show that combining the pixel-wise loss with structural or perceptual terms can affect the reconstruction quality, revealing a tradeoff between the perceptual quality and numerical accuracy in restoration tasks.

4.4 | Comparisons with existing models

Here, we discuss both quantitative and qualitative evaluations to compare the performance of the proposed model with those of the existing generative methods. Table 4 presents a quantitative evaluation of the image restoration quality for each model using the four representative metrics.

The RICH model outperformed all the others by achieving the lowest MSE and LPIPS values as well as the highest PSNR and MS-SSIM scores, indicating superior pixel-level accuracy, perceptual similarity, and structural consistency. The Context encoder [44] exhibited the weakest performance, with high MSE and LPIPS values and the lowest MS-SSIM score. The SDXL model [45] achieved lower MSE and LPIPS values than those of the Context encoder, and its MS-SSIM score improved enormously but still fell short of those of the FLUX and RICH models. The FLUX model [46] further improved the PSNR while maintaining a reasonable LPIPS value;

TABLE 3 Quantitative comparison of different loss function.

Loss	Metric			
	MSE ↓	PSNR ↑	LPIPS ↓	MS-SSIM ↑
MAE + SSIM	51.6350	10.2030	0.1506	0.9445
MSE + SSIM	**33.3374**	**29.8862**	**0.1141**	**0.9761**
MAE + LPIPS	36.8720	11.4630	0.0997	0.9545
MSE + LPIPS	46.0146	10.5933	0.1036	0.9498

TABLE 4 Quantitative evaluation of restoration quality for different models.

Model	Metric			
	MSE ↓	PSNR ↑	LPIPS ↓	MS-SSIM ↑
Context encoder [44]	379.7840	22.4480	0.6610	0.7997
SDXL [45]	300.3232	23.6859	0.1506	0.9445
FLUX [46]	243.2862	24.5260	0.1985	0.9284
RICH	33.3374	29.8862	0.1141	0.9761

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

however, it remained less effective than the RICH model in terms of both pixel-level accuracy and structural similarity. These results clearly demonstrate the superior restoration performance of the RICH model compared to the other baseline models.

As shown in Figure 5, the proposed model demonstrated powerful Chinese inscription generation capabilities with clear results. By contrast, the Context encoder generated restored areas with square-shaped noise, failing to properly generate Chinese inscriptions on the surface and incorrectly detecting restoration areas in incorrect locations, owing to poor recognition. The SDXL and FLUX models allowed users to specify the restoration areas using masks, with the results being generated using the prompt "restore the inscriptions."

The SDXL output images were more blurred than the input images, with some masked areas either completely unrestored or generating unclear imitations of the Chinese inscriptions. The FLUX model also produced unclear results and, in some cases, generated English words.

Consequently, we conducted a user study to evaluate the restoration performance of the proposed model based on the images shown in Figure 5 and other related samples. Figure 6 shows part of the questionnaire used during the study. In the user study, participants were asked to rank the images generated by the four different models based on the perceived quality of restoration. The participants included graduate students from the Department of Artificial Intelligence, Department of Digital Cultural Heritage, and working professionals with field experience. The ranking results were aggregated and analyzed using the Borda Count [47] method, providing a fair and robust comparison across the models.

Based on the results summarized in Figure 7, the RICH model consistently received the highest rankings across practically all questions, demonstrating a clear preference for its restoration results among the participants. The SDXL and FLUX models exhibited moderate performances, whereas the Context encoder ranked the lowest in most cases.

This trend was further reflected in the overall scores summarized in Table 5, where the RICH model



FIGURE 5 Sangpyeongtongbo qualitative results. Comparison with other generative models.

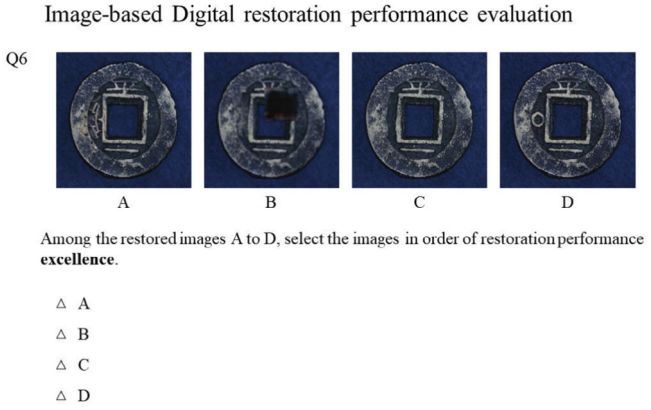


FIGURE 6 A section of the questionnaire used in the user study.



FIGURE 7 Ranking results from the user study using the Borda Count method per question.

TABLE 5 Overall Borda Count scores aggregated per model and normalized to a 100-point scale based on user study rankings.

Model	Context encoder	SDXL	FLUX	RICH
Score	31.9	78.2	68.5	97.4

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

outperformed the other models, with a normalized Borda Count score of 97.4/100. These findings suggest that the RICH model not only excelled in quantitative metrics but also delivered superior perceived quality in human evaluations, confirming its practical effectiveness in real-world restoration scenarios.

Accordingly, we could confirm that the proposed model excelled in restoring damaged Chinese inscriptions to their original form. The superior performance of the proposed model can be attributed to several key factors. First, it maintained consistency in script style and scale across the entire restoration area, preserving the aesthetic uniformity of the original coin surface. Additionally, the model demonstrated remarkable robustness against various types of degradation patterns, including scratches, wear, and corrosion, consistently producing accurate restorations regardless of the damage type. Moreover, unlike other methods that require multiple iterations or manual adjustments, the proposed model achieved high-quality results in a single pass, making it more efficient and practical for numerical research and conservation. Notably, even with high-resolution images, the model accurately reproduced the intricate strokes and structural features of inscriptions, achieving results of sufficient quality to serve as reference material for actual coin identification. This performance difference demonstrated that the RICH model excelled in effectively understanding the complex structures of damaged images and restoring them to their original state, more so than the other methods.

5 | CONCLUSIONS

In this study, we introduced the RICH model, an algorithm designed for the digital restoration of ancient Korean coins. The key idea was to use diffusion networks composed of a U-Net and VAE to restore the damaged information on the coin surfaces to their original form. Specifically, we developed a diffusion-model-based RICH model to fully reconstruct Chinese inscriptions engraved on coin surfaces. The RICH model, trained on the Sangpyeongtongbo dataset, effectively restored Chinese inscriptions on coins that were difficult to decipher because of corrosion and damage, thereby providing and identifying crucial information about the coins. The proposed approach exhibited several major advantages: it maintained consistent restoration quality across various degradation patterns, achieved high accuracy in inscription reconstruction without compromising the aesthetic integrity of the original artifacts, and provided reliable results in a single processing step. We demonstrated that the RICH model could not only outperform existing restoration methods but could also offer a practical and efficient solution for cultural heritage preservation. Moreover, it could provide practical assistance to cultural heritage experts in reconstructing the original appearance of damaged artifacts, while also demonstrating the potential use of generative models in the digital restoration of damaged cultural properties. Finally, this work establishes a promising framework for extending restoration methods to other types of historical artifacts, potentially revolutionizing the field of digital cultural heritage preservation.

ACKNOWLEDGMENTS

This work was partly supported by the Artificial Intelligence Industrial Convergence Cluster Development Project funded by the Ministry of Science and ICT (MSIT, Republic of Korea) & Gwangju Metropolitan City, Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: RS-2023-00219579), and Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2021-II211341, Artificial Intelligence Graduate School Program, Chung-Ang University).

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

ORCID

Seohyun Baek https://orcid.org/0009-0008-4590-5475
Jae-Ho Lee https://orcid.org/0000-0001-6773-9350
Sangdoo Ahn https://orcid.org/0000-0003-3803-9210
Jongwon Choi https://orcid.org/0000-0001-9753-8760

REFERENCES

1. Ministry of Government Legislation, Korean Law Information Center, *Act on the Preservation and Utilization of Cultural Heritage*, 2024. https://www.law.go.kr/LSW/lsInfoP.do?urlMode=lsInfoP&lsId=001607 [last accessed November 2024]
2. S. J. Lee, *Preliminary study on defining and assessing heritage values for establishing conservation principles*, Korean J. Heritage: History Sci. **44** (2011), no. 4, 154–171.
3. S. Nam and S. Jo, *Study of historical developments and design of money*, (Proc. Conf. Korean Contents Society Spring Academic Conf., Daejeon, Republic of Korea), 2017, pp. 163–164.
4. S. Baek, *Classification and digital restoration of Korean coins using deep learning and reflectance transformation imaging*, Ph.D. Dissertation, Cultural Properties Department, Chung-Ang University, Seoul, Republic of Korea, 2024.
5. J. H. Kim and G. H. Kim, *A study on the physical and chemical characteristics of Sangpyeongtongbo according to the casting period and mint in Joseon dynasty*, Sci. Eng. Cult. Herit. **17** (2022), no. 1, 63–74.
6. S. B. Jang, N. C. Cho, and H. T. Kang, *Study on the materials characteristics of Sangpyeongtongbo coins in Joseon dynasty using chemical compositions and microstructures*, Conserv. Sci. **31** (2015), no. 3, 319–330.

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7. S. Baek, Y. Cho, S. Ahn, and J. Choi, *Algorithm development for texture and color style transfer of cultural heritage images*, Conserv. Sci. Museum **31** (2024), no. 4, 55–70.
8. A. Koutsoudis, B. Vidmar, and F. Arnaoutoglou, *Performance evaluation of a multi-image 3D reconstruction software on a low-feature artefact*, J. Archaeol. Sci. **40** (2013), no. 12, 4450–4456.
9. F. Bruno, S. Bruno, G. De Sensi, M.-L. Luchi, S. Mancuso, and M. Muzzupappa, *From 3D reconstruction to virtual reality: A complete methodology for digital archaeological exhibition*, J. Cult. Herit. **11** (2010), no. 1, 42–49.
10. F. Kleber and R. Sablatnig, *A survey of techniques for document and archaeology artefact reconstruction*, (Proc. IEEE International Conference on Document Analysis and Recognition, Barcelona, Spain), 2009, pp. 1061–1065.
11. A. Aggarwal, M. Mittal, and G. Battineni, *Generative adversarial network: An overview of theory and applications*, Int. J. Inf. Manag. Data Insights **1** (2021), no. 1, 100004.
12. M. Altaweel, A. Khelifi, and M. H. Zafar, *Using generative AI for reconstructing cultural artifacts: Examples using Roman coins*, J. Comput. Appl. Archaeol. **7** (2024), no. 1, 301–315.
13. M. Zachariou, N. Dimitriou, and O. Arandjelović, *Visual reconstruction of ancient coins using cycle-consistent generative adversarial networks*, Sci. **2** (2020), no. 3, 52.
14. O. Elharrouss, R. Damseh, A. N. Belkacem, E. Badidi, and A. Lakas, *Transformer-based image and video inpainting: Current challenges and future directions*, Artif. Intell. Rev. **58** (2025), no. 4, 1–45.
15. T. Ning, G. Huang, J. Li, and S. Huang, *Complex image inpainting of cultural relics integrating multi-stage structural features and spatial textures*, Pattern Anal. Appl. **28** (2025), no. 2, 1–18.
16. Z. Guo, H. Shao, X. Deng, and Y. Jiang, *DenoSR: A high-fidelity super-resolution approach for noisy images*, ETRI J. (2025), 1–11. DOI 10.4218/etrij.2024-0295
17. M. Wang, J. Yan, C. Wan, G. Yang, and T. Yu, *A bi-stream transformer for single-image dehazing*, ETRI J. **47** (2024), 545–558.
18. Z. Wang, J. Jia, and J. Min, *Multiscale implicit frequency selective network for single-image dehazing*, ETRI J. (2024), 1–12. DOI 10.4218/etrij.2024-0146
19. M. Elgendy and H. Sim, *Deep learning for vision systems*, Hanbit Publishing Network, Seoul, Republic of Korea, 2022, 409–415.
20. Stanford University, *CS231n: Deep learning for computer vision*, 2023. https://cs231n.stanford.edu/2023/index.html [last accessed Jun 2023]
21. D. P. Kingma, *Auto-encoding variational bayes*, arXiv Preprint, (2013). DOI 10.48550/arXiv.1312.6114
22. A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, *Repaint: Inpainting using denoising diffusion probabilistic models*, (Proc. Conf. IEEE/CVF conference on computer vision and pattern recognition, New Orleans, LA, USA), June 2022, pp. 11461–11471.
23. A. Yukinaga, J. Choi, and G. Jang, *Introduction to core deep learning: learn only the latest deep learning technologies*, only-book, Anyang, Republic of Korea, 2020, 262–332.
24. L. Zhang, Y. Yu, J. Yao, and H. Fan, *High-fidelity image inpainting with multimodal guided GAN inversion*, Int. J. Comput. Vis. **133** (2025), no. 8, 1–5805.
25. W. Lu, H. Zhao, X. Jiang, X. Jin, Y. Yang, M. Wang, J. Lyu, and K. Shi, *Do inpainting yourself: Generative facial inpainting guided by exemplars*, Neurocomput. **617** (2025), 128996.
26. Y. Yun, N. Cho, Y. Jeong, and I. Lim, Study on the casting technology and restoration of "Sangpyong Tongbo,", MUNHWAJAE Korean J. Cult. Herit. Studie **47** (2014), no. 4, 224–243.
27. I. Dumitriu, R. C. Fierascu, A. Catangiu, M. Neata, R. M. Ion, and R. Somoghi, *A synthetic approach into the restoration and conservation of metal artifacts (coins)*, J. Optoelectron. Adv. Mater. **13** (2011), no. 7, 874–878.
28. C. Klose, *Seriality and restoration: The 'Restored Coins' of the roman empire*, (Proc. Conf. 19th International Congress of Classical Archaeology, Cologne, Germany), May. 2018, pp. 31–43.
29. B. Marshall, *'Restoration' coins of the flavian emperors, Vespasian and Titus*, J. Numismatic Assoc. Aust. **32** (2023), no. 6, 91–104.
30. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, *Image-to-image translation with conditional adversarial networks*, (Proc. Conf. IEEE Conference on Computer Vision and Pattern Recognition, Venice, Italy), 2017, pp. 1125–1134.
31. S. Baek, H. Hwang, C. W. Park, H. K. Kim, J. H. Lee, S. Ahn, and J. Choi, *Proposal for digital restoration of ancient Korean coins using images of Sangpyeongtongbo*, (Proc. Conf. The 1st International Conference on Artificial Intelligence Computing and Systems, Jeju, Republic of Korea), 2024, pp. 102–106.
32. B. A. Lietaer and N. Kang, *Money, its soul and truth: in search of the essence and history of money*, Chamsol, Seoul, Republic of Korea, 2004, 126–131.
33. C. Lee, *Why was the sangpyeong notice created*, Hist. That Opens Tomorrow **9** (2002), no. 9, 133–144.
34. A. Sargsyan, S. Navasardyan, X. Xu, and H. Shi, *Mi-gan: A simple baseline for image inpainting on mobile devices*, (Proc. Conf. IEEE/CVF International Conference on Computer Vision, Paris, France), Oct. 2023, pp. 7335–7345.
35. J. Ho, J. Ajay, and P. Abbeel, *Denoising diffusion probabilistic models*, arXiv Preprint, (2020). DOI 10.48550/arXiv.2006.11239
36. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, *High-resolution image synthesis with latent diffusion models*, arXiv Preprint, (2021). DOI 10.48550/arXiv.2112.10752
37. O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, (Proc. Med. Image Comput. Comput-Assist. Interv., Munich, Germany), Oct. 2015, pp. 231–241.
38. E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, *LoRA: Low-rank adaptation of large language models*, arXiv Preprint, (2021). DOI 10.48550/arXiv.2106.09685
39. J. Ren, M. Zhang, C. Yu, Z. Liu, *Balanced MSE for imbalanced visual regression*, (Proc. IEEE/CVF conf. Comput. Vis. Pattern Recognit., New Orleans, LA, USA), June 2022, pp. 7916–7925.
40. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Trans. Image Process. **13** (2004), no. 4, 600c612.
41. K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, *Masked autoencoders are scalable vision learners*, (Proc. IEEE/CVF conference on computer vision and pattern recognition, New Orleans, LA, USA), June 2022, pp. 16000–16009.

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

42. A. Hore and D. Ziou, *Image Quality Metrics: PSNR vs. SSIM*, (Proc. 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey), Aug. 2010, pp. 2366–2369.
43. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, *The unreasonable effectiveness of deep features as a perceptual metric*, (Proc. IEEE Conference on Computer vision and Pattern Recognition, Salt Lake City, UT, USA), 2018, pp. 586–595.
44. D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, *Context encoders: Feature learning by inpainting*, (Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Las Vegas, NV, USA), June 2022, pp. 2536–2544.
45. F. R. Perche-Mahlow, A. Felipe-Zanella, W. A. Cruz-Castañeda, and M. Amadeus, *An inpainting-infused pipeline for attire and background replacement*, arXiv Preprint, (2014). DOI 10.48550/arXiv.2402.03501
46. J. Plin, *FLUX-Controlnet-Inpainting*, 2024. https://github.com/alimama-creative/FLUX-Controlnet-Inpainting [last accessed November 2024]
47. G. W. Bassett Jr. and J. Persky, *Robust voting*, Public Choice **99** (1999), no. 3, 299–310.

AUTHOR BIOGRAPHIES



Seohyun Baek received her BFA degree in ceramic arts and design from the Seoul National University of Science and Technology, Seoul, Republic of Korea, in 2018; MS degree in conservation of cultural heritage from the Graduate School of Yong In University, Yongin, Republic of Korea, in 2020; and PhD degree in science from the Chung-Ang University, Seoul, Republic of Korea, in 2024. She is currently a post-doctoral researcher at the Hyper-Reality Metaverse Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea. Her research interests include AI-based cultural heritage restoration, digital heritage, cultural heritage analysis, and conservation.



Jae-Ho Lee received his PhD in electronic engineering from Hanyang University, Seoul, Republic of Korea, in 2005, and BS and MS degrees in electronic engineering from the same university in 1999 and 2001, respectively. He has been working at the Electronics and Telecommunications Research Institute since 2005. His main research interests include face recognition, multimedia management, and artificial intelligence applications related to cultural heritage content.



Sangdoo Ahn received his PhD in chemistry from Seoul National University, Republic of Korea in 1995. He has been a professor of chemistry at Chung-Ang University since 2004. His research focuses on developing analytical methods for various substances in industrial materials and foods using analytical instruments, such as NMR. He also serves as an adjunct professor in the Department of Cultural Properties, where he conducts scientific analyses of various cultural properties. Recently, his work incorporated artificial intelligence techniques for data analysis.



Jongwon Choi received his BS and MS degrees in electrical engineering from KAIST, Daejeon, Republic of Korea, in 2012 and 2014, respectively, and PhD degree in electrical engineering from Seoul National University, Seoul, Republic of Korea, in 2018. From 2018 to 2020, he worked as a research engineer at the Research Intelligence Research Center, Samsung SDS, Seoul. In 2020, he joined the Department of Advanced Imaging, Chung-Ang University, Seoul, where he is currently working as an Associate Professor. His research interests include deep learning, computer vision, and artificial intelligence safety.

How to cite this article: S. Baek, J.-H. Lee, S. Ahn, and J. Choi, *RICH: Coin data restoration using inpainting for cultural heritage analysis*, ETRI Journal **48** (2026), 464–475, DOI 10.4218/etrij.2025-0113

22337326, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2025-0113 by Chung-Ang University, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License