The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W18-2025
Symposium on GeoSpatial Technologies: Visions and Horizons 2025 (GeoVisions2025), 8–10 October 2025, Çanakkale, Türkiye

Generative Inpainting of Partially destroyed Frescos

Vladimir V. Kniaz[1,2], Petr V. Moskantsev[2], Artem N. Bordodymov[2], Vladimir A. Knyaz[1,2], Vladislav Pashtanov[2], Aleksandr Dezhin[2], Tatyana N. Skrypitsyna[3], Victor Aleksandrov[2]

[1] Moscow Institute of Physics and Technology (MIPT), Moscow, Russia - (kniaz.vv, kniaz.va)@mipt.ru
[2] State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia - zhl@gosniias.ru
[3] Moscow State University of Geodesy and Cartography (MIIGAiK), Moscow, Russia - tatyana.skrypitsyna@yandex.ru

Keywords: cultural heritage, neural networks, diffusive network model, generative adversarial learning.

Abstract

Fresco painting, a prevalent mural technique, is highly susceptible to damage due to the hygroscopic nature of lime plaster, often resulting in partially destroyed artworks. Restoring these masterpieces poses significant challenges: merging remaining fragments with new additions requires skilled restoration, while reconstructing original compositions demands expert historical insight. The absence of textual or graphical records further complicates this task. Recent advancements in neural inpainting offer potential solutions but lack the precision required by art historians. We introduce the Stable Restorer model, enhancing neural inpainting with fine-grained control over generated content. Building on the Flux model with LoRA adaptation, our approach employs multiple text prompts linked to attention masks for precise local editing. Our Mural Paintings dataset, comprising 15k samples of diverse frescos, facilitates rigorous evaluation. Results indicate that our model not only competes with but surpasses state-of-the-art methods, offering art historians enhanced control over reconstructed fresco regions.

1. Introduction

The restoration of partially destroyed frescos represents a crucial intersection of art preservation and technological innovation. Frescos, as a significant cultural heritage, embody historical narratives and artistic achievements from various civilizations. However, due to their inherent fragility and the hygroscopic nature of lime plaster, many frescos have suffered damage over time, resulting in partial destruction. Traditional restoration methods demand not only artistic skill but also historical expertise to accurately reconstruct these works. This task is further complicated by the absence of comprehensive records or descriptions for many frescos. In recent years, generative neural networks have emerged as promising tools for digital inpainting, offering new possibilities for the restoration of these invaluable artworks. By leveraging advanced machine learning techniques, it is possible to reconstruct missing sections with remarkable fidelity, thereby preserving the integrity and continuity of the original masterpiece.

The importance of restoring partially destroyed frescos extends beyond immediate aesthetic recovery; it plays a pivotal role in safeguarding cultural heritage for future generations. Accurate digital restorations serve as essential references for ongoing and future physical restoration efforts, ensuring that any interventions remain true to the original artist's vision. Moreover, creating detailed digital documentation of restored frescos is vital in the event of further deterioration or catastrophic loss due to environmental factors or human activities. Such documentation not only aids in potential future reconstructions but also serves as an educational resource that can be shared globally, promoting cultural appreciation and scholarly research. Thus, integrating generative neural networks into fresco restoration not only enhances current preservation practices but also fortifies our ability to protect and celebrate our shared artistic legacy against unforeseen challenges.

The current state-of-the-art in the restoration of ancient relics, including frescos, is characterized by the use of advanced deep learning algorithms that facilitate image color restoration. A notable advancement in this field was introduced in 2021, leveraging the DenseNet model (Xu and Fu, 2023) to create an interactive digital mural-restoration system. This approach has demonstrated significant efficacy in reconstructing and enhancing the color fidelity of ancient artworks, offering a powerful tool for art conservators. However, despite these advancements, a critical challenge remains: achieving a flexible and precise inpainting of missing sections in partially destroyed frescos. Current models often lack the capability for fine-grade control over reconstructed details, which is essential for maintaining historical accuracy and artistic integrity. Addressing this issue requires developing methods that allow for nuanced control over the generated content, enabling restorers to specify exact locations and characteristics of reconstructed regions. This paper aims to resolve these limitations by introducing a novel approach that integrates multiple text prompts with attention masks, thereby enhancing the precision and flexibility of fresco restoration efforts.

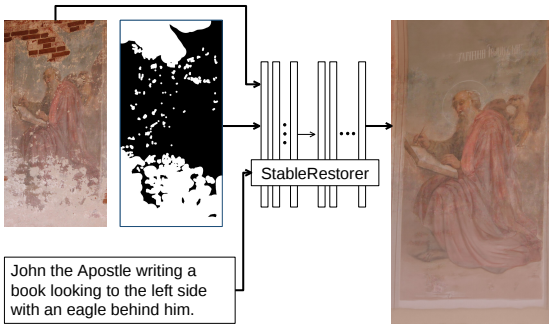

Figure 1. Restoration of partially destroyed fresco using our proposed `StableRestorer` model.

The authors of this paper are engaged in a comprehensive research project dedicated to advancing methodologies for the documentation and digital restoration of partially destroyed cul-

tural heritage artifacts. A significant milestone within this project has been the development of detailed 3D models of two historic churches, achieved through photogrammetry and the structure-from-motion approach. While these techniques have become indispensable tools for researchers seeking to capture and preserve the geometric intricacies of architectural heritage, there remains a pressing need for innovative methods that facilitate human-guided completion of missing fresco elements. The objective of this study is to bridge this gap by developing a flexible inpainting framework that empowers restorers with fine-grade control over the reconstruction process. By integrating advanced generative neural network techniques with user input, the authors aim to enhance the precision and authenticity of digital fresco restorations, thereby contributing to the broader field of cultural heritage preservation.

The restoration and preservation of partially damaged objects of cultural heritage have long been a priority for art historians and conservationists, driven by the need to maintain the historical and aesthetic value of these artifacts for future generations. The task of reconstructing frescos with missing fragments has garnered considerable attention in recent literature, reflecting the growing demand for innovative solutions in this domain. Traditional methods of restoration, while effective to some extent, often fall short in addressing the complexities involved in accurately reconstructing intricate details lost over time. This has led to an increasing interest in leveraging modern computational techniques to enhance restoration efforts.

In 2022, a significant advancement was made with the introduction of a deep learning algorithm specifically designed for the restoration of ancient relics (Xu and Fu, 2023). This algorithm utilizes a DenseNet model, which incorporates residual blocks to facilitate a detailed analysis of mural paintings, thereby improving color fidelity and structural integrity in restored images. Building on this foundation, 2023 saw the proposal of employing generative adversarial networks (GANs) for the restoration of damaged artworks (Kumar and Gupta, 2023). GANs have shown promise in generating high-quality reconstructions by learning complex patterns and textures inherent in artistic works. Despite these advancements, there remains a critical need for methods that offer more nuanced control over the reconstruction process, allowing restorers to guide the completion of missing fresco parts with precision. This paper aims to address this gap by developing a flexible generative inpainting approach that integrates user input for enhanced fine-grade control over reconstructed details.

Recent advancements in generative models have introduced a new paradigm known as diffusion models, which have demonstrated remarkable capabilities in image synthesis and inpainting tasks (Sohl-Dickstein et al., 2015, Ho et al., 2020). Notable implementations of these models, such as StableDiffusion and DALLE-2, have shown proficiency in filling missing parts of images guided by natural language instructions (Rombach et al., 2022, Ramesh et al., 2022). This capability is particularly relevant to the restoration of missing parts in partially destroyed frescos, where the integration of semantic guidance can significantly enhance restoration accuracy. The application of diffusion models to the reconstruction of destroyed art objects has seen a surge in research interest. In 2023, a study utilized a diffusion model to effectively repair cracks in mural paintings, showcasing its potential for detailed restoration work (Huang and Hong, 2023). Following this, the development of the Flux-1 diffusion model in 2024 further improved the performance of text-guided image inpainting, offering more precise control over restored images (Black Forest Labs et al., 2025). These advancements underscore the transformative potential of diffusion models in cultural heritage preservation and motivate further exploration into their application for fresco restoration.

To the best of our knowledge, there has been no research to date that specifically addresses the use of multimodal-guided models for the restoration of partially destroyed frescos. In this study, we leverage the Flux Kontext diffusion model as a foundational framework to explore this novel approach. Our primary aim is to train the Flux Kontext model to effectively inpaint missing regions of frescos by utilizing a dual-guidance system: a semantic scheme provided by an art historian and a traditional textual prompt. This innovative approach allows for a more contextually aware restoration process, integrating expert historical insights with advanced computational techniques. To facilitate this research, we have generated a synthetic dataset comprising pairs of artificially degraded frescos, their corresponding semantic segmentation schemes, and their original untouched versions. This dataset serves as a crucial resource for training and evaluating the model's performance in accurately reconstructing lost fresco details while adhering to historical authenticity. Through this work, we aim to establish a new paradigm in digital fresco restoration that bridges the gap between art historical expertise and state-of-the-art generative modeling.

The aim of the present work is to develop a diffusion neural model specifically designed for the reconstruction of missing parts in partially destroyed frescos. This model seeks to integrate multiple forms of input data to enhance the accuracy and authenticity of the restoration process. Specifically, the model will take as input an image of the partially destroyed fresco, a semantic sketch of the missing part provided by an art historian, and a textual prompt that describes the content and context of the missing sections. By combining these diverse inputs, our approach aims to leverage both visual and semantic information, thereby facilitating a more informed and nuanced restoration process. This multimodal integration is intended to capture not only the aesthetic qualities but also the historical and cultural significance embedded within the original artwork. Through this work, we aspire to contribute a robust tool for cultural heritage preservation, offering a sophisticated method for digitally restoring frescos with a level of detail and contextual awareness that respects their historical integrity.

The results of our study are encouraging and illustrate that our approach not only achieves but also competes with the state-of-the-art in the generative reconstruction of damaged artistic pieces. Specifically, we have developed a model named StableRestorer, which was rigorously compared against three modern generative models. Utilizing the test portion of the SemanticFresco dataset, we evaluated both our model and the baseline models on their ability to complete the missing parts of frescos. Given that the artificially destroyed images in this dataset have corresponding untouched versions, these serve as ground truth for validation purposes. Our evaluation employed two key metrics: reverse mean Average Precision (r-mAP) and Fréchet Inception Distance (FID). The results demonstrate that our approach outperforms the baseline models by a significant margin—achieving a 15% improvement in the r-mAP metric and a 21-point reduction in FID. These findings underscore the efficacy of our method in producing high-quality reconstructions that align closely with the original artistic intent, thereby offering a promising advancement in digital fresco restoration methodologies.

The future implications of our StableRestorer model are manifold, offering significant advancements in both academic and practical domains. For art historians, the model serves as a rapid prototyping tool, enabling the generation of hypothetical reconstructions of partially destroyed frescos with unprecedented speed and accuracy. This capability allows historians to explore various restoration scenarios and gain deeper insights into the potential original appearances of these artworks. Additionally, the model holds promise for development into an interactive software platform, empowering users to engage directly with the reconstruction process. Such a platform could enhance public understanding and appreciation of cultural heritage preservation by emphasizing the critical importance of physical restoration efforts for art pieces at risk of total destruction. By facilitating user interaction and experimentation, this application could also foster greater awareness and support for conservation initiatives, thereby contributing to the safeguarding of artistic heritage for future generations.

2. Related Work

2.1 Image Inpainting

The field of image inpainting has undergone significant evolution over the past few decades, transitioning from traditional statistical methods to advanced deep learning techniques. Early approaches to image inpainting, such as the work by Ružić and Pižurica (Ružić and Pižurica, 2015), employed Markov Random Fields (MRFs) to model and reconstruct missing parts of images. However, with the advent of generative neural models in 2014, deep learning rapidly became the standard tool for tackling this problem. The introduction of Generative Adversarial Networks (GANs) marked a pivotal moment in this transition. Ian Goodfellow's seminal work on unconditional GANs (Goodfellow et al., 2014) laid the foundation for using adversarial training to generate realistic images. Building on this, conditional GANs were introduced by Isola et al. (Isola et al., 2017) in 2016, allowing for more controlled and context-aware image generation. Further advancements were made by Zhu et al. (Zhu et al., 2017) in 2017, improving the quality and applicability of GANs for various image-to-image translation tasks. Despite these advancements, GAN-based methods often suffer from artifacts such as checkerboard patterns and random color blobs, which can detract from the photorealism of the generated images. Recently, diffusion models have emerged as a promising alternative, offering enhanced inpainting quality by addressing some of these limitations inherent in GAN-based approaches.

2.2 Diffusion Models

The advent of diffusion generative models has marked a significant leap forward in the field of image inpainting, offering improved quality and robustness over previous methods. Diffusion models operate by iteratively refining a noisy image until it converges to a coherent and realistic output. One of the pioneering works in this domain was introduced by Sohl-Dickstein et al. (Sohl-Dickstein et al., 2015), who proposed a framework for deep unsupervised learning using diffusion processes. This foundational work paved the way for subsequent advancements, including the development of more sophisticated architectures such as Denoising Diffusion Probabilistic Models (DDPM) by Ho et al. (Ho et al., 2020), which demonstrated impressive capabilities in generating high-fidelity images.

Among the notable diffusion models are Stable Diffusion (Rombach et al., 2022) and DALL-E 2 (Ramesh et al., 2022), both of which have garnered significant attention for their ability to generate detailed and contextually appropriate images from textual descriptions. These models leverage large-scale training datasets and advanced neural architectures to achieve state-of-the-art results in various generative tasks, including image inpainting.

Additionally, recent innovations have introduced models like Flux Kontext (Black Forest Labs et al., 2025), which further enhance the contextual understanding and generation capabilities of diffusion models. Other noteworthy contributions include Score-Based Generative Models by Song et al. (Song and Ermon, 2019) and Improved DDPMs by Nichol and Dhariwal (Nichol and Dhariwal, 2021), each contributing unique insights into the optimization and scalability of diffusion-based approaches.

These advancements collectively underscore the potential of diffusion generative models to transform image inpainting, offering tools that are not only capable of producing visually compelling results but also adaptable to a wide range of artistic and practical applications (Avena et al., 2024, Kniaz et al., 2024b, Kniaz et al., 2024a).

2.2.1 Generative Restoration of Cultural Heritage The generative restoration of cultural heritage artifacts, particularly damaged frescos, has become an increasingly prominent area of research within the field of computer vision. This interest is driven by the potential to digitally reconstruct and preserve invaluable historical objects that have suffered deterioration over time (Zhao et al., 2025, Knyaz et al., 2024, Kniaz et al., 2025). Since the inception of computer vision, researchers have been exploring methods to aid in the digital restoration of these cultural treasures. In 2024, Zhao and Ren introduced a GAN-based heterogeneous network specifically designed for the restoration of ancient murals (Zhao et al., 2025). Their innovative approach combined convolutional networks with transformers to enhance reconstruction quality, marking a significant advancement in the application of generative models to cultural heritage. Building on this work, Zhao and Ren further advanced their methodology in 2025 by leveraging diffusion models for mural reconstruction, which demonstrated improved fidelity in restoring intricate details of damaged artworks. Despite the power of generative models in achieving photorealistic inpainting of missing regions, ensuring that reconstructed fragments are coherent with authentic lost pieces remains a challenging issue. It is evident that human-guided models can produce outputs that are more historically accurate than those trained solely on perceptual image quality metrics. Addressing this challenge, Hu and Yu proposed the GuidePaint image-guided diffusion model in 2025 (Hu et al., 2025), introducing an image sampling algorithm that facilitates an unsupervised training pipeline. This approach underscores the importance of integrating human expertise into generative restoration processes to achieve more historically sound reconstructions.

3. Method

The objective of this study is to develop a novel generative model, termed StableRestorer, designed for the human-guided restoration of partially destroyed frescos. The StableRestorer framework is engineered to integrate multiple inputs to achieve high-fidelity reconstructions that are both visually coherent and

historically authentic. Specifically, our model accepts a color image of the damaged fresco as its primary input, supplemented by an additional hint input that provides the semantic structure of the missing fragments. This hint input is crucial for guiding the model in understanding spatial relationships and contextual cues within the fresco. Furthermore, a textual description of the authentic fresco is incorporated to ensure that the generated output aligns with historical and artistic expectations. By synthesizing these diverse forms of input, StableRestorer aims to produce restorations that not only fill in missing regions but also respect the original artistic intent and cultural significance.

The remainder of this section is structured to provide a comprehensive overview of our approach. We begin by outlining the StableRestorer framework, detailing its core components and workflow. Following this, we delve into the architectural design of our model, elucidating how it integrates convolutional networks with transformer mechanisms to enhance reconstruction quality. We also introduce our proposed loss function, which balances perceptual image quality with semantic coherence. Lastly, we discuss our strategy for generating the SemanticFresco dataset, which serves as a critical resource for training and validating our model. This dataset is meticulously curated to include a wide array of fresco styles and conditions, ensuring that StableRestorer is equipped to handle diverse restoration scenarios effectively.

3.1 Framework Overview

The StableRestorer framework is structured around four distinct domains to facilitate the restoration of partially destroyed frescos. These domains include the damaged fresco domain $\mathcal{A} \in \mathbb{R}^{w \times h \times 3}$, which represents the RGB image of the fresco with dimensions w (width) and h (height); the semantic annotation domain $\mathcal{S} \in \{0, 1, \ldots, k\}^{w \times h}$, where each pixel is assigned a label from k semantic classes; the restored fresco domain $\mathcal{B} \in \mathbb{R}^{w \times h \times 3}$, representing the target output; and the textual prompt domain $\mathcal{T}$, which provides descriptive guidance for restoration. Our primary objective is to train a generative mapping function $G : (A, S, T) \rightarrow B$, which takes as input a damaged fresco A, its corresponding semantic annotation S, and a textual description T, and outputs a restored fresco B. This approach builds upon and extends the capabilities of existing models by incorporating semantic structure into the restoration process.

To achieve this integration, we utilize the Flux Kontext model as our foundational architecture. The original Flux Kontext model facilitates a mapping $F : (A, T) \rightarrow B$, leveraging both visual and textual inputs for restoration. To incorporate the additional semantic input S, we adapt the encoder network of the Flux Kontext model. This encoder processes the input image to generate a latent feature map x_A. Similarly, we encode the semantic annotation S into another latent feature map x_S. These two feature maps are then concatenated to form a comprehensive latent representation that captures both visual and semantic information. Mathematically, this can be expressed as:

$$x_{\text{combined}} = [x_A; x_S]$$

where $[;]$ denotes concatenation along the appropriate dimension. This enriched latent representation is subsequently fed into subsequent layers of our network to produce high-quality restorations that are semantically informed and visually coherent. By integrating these diverse data modalities, StableRestorer aims to achieve superior restoration outcomes that honor both artistic integrity and historical accuracy.

3.2 Network Architecture

The architecture of the StableRestorer diffusion neural network builds upon the foundational principles of the Flux Kontext model, which excels in contextual image generation and editing. The Flux Kontext model is uniquely designed to integrate an existing image with a textual instruction, producing a new image that adheres to the given instruction while maintaining the original's style and structural integrity. This capability distinguishes it from conventional image generation models, which often treat each new prompt as an independent task without regard for context. By understanding and preserving context, Flux Kontext can modify elements such as characters without altering their identity, maintain consistent layouts, and uphold a visual style across multiple generations.

At the core of the Flux Kontext model are two pivotal concepts: flow matching and latent space editing. Flow matching is a sophisticated technique for training generative models by learning to reverse a noise process through predicting a continuous "flow" that transitions data from a noisy state back to its clean form. Unlike traditional diffusion methods that rely on step-by-step noise reversal, flow matching operates in latent space—a compressed representation of images—allowing for efficient manipulation of complex data structures. In this latent space, the model learns mappings between noisy and clean representations, guided by contextual inputs such as an initial image and corresponding natural-language instructions. For text-to-image tasks, Flux Kontext begins with random noise and progresses toward generating an image. In contrast, for image-to-image editing tasks like ours, it starts with a real image, encodes it into latent space, and transforms it based on the edit prompt.

Our StableRestorer model introduces two significant enhancements to the original Flux Kontext framework. Firstly, we encode both the damaged fresco image and its semantic annotation input into separate latent representations, which are then concatenated and fused into a single comprehensive latent vector. This integration ensures that both visual details and semantic structures are considered during restoration. Secondly, we implement a modified semantic loss function, detailed in the subsequent section, which further refines our model's ability to produce semantically coherent restorations. The architecture of our StableRestorer model is illustrated in Figure 2, showcasing these modifications and their integration within the broader framework. By building upon these advanced techniques, StableRestorer aims to deliver restorations that are not only visually accurate but also contextually faithful to historical artistry.

3.3 Loss Function

In the training of our StableRestorer model, we employ a sophisticated loss function that leverages classifier-free guidance to enhance both the photorealism of the output images and their coherence with semantic inputs. This approach is inspired by the findings of (Wang et al., 2023), who demonstrated that classifier-free guidance can significantly improve image generation quality in terms of visual fidelity and semantic alignment. The essence of this technique lies in its ability to guide the generative process without relying on explicit class labels, thus allowing for more flexible and nuanced modeling of complex data distributions.

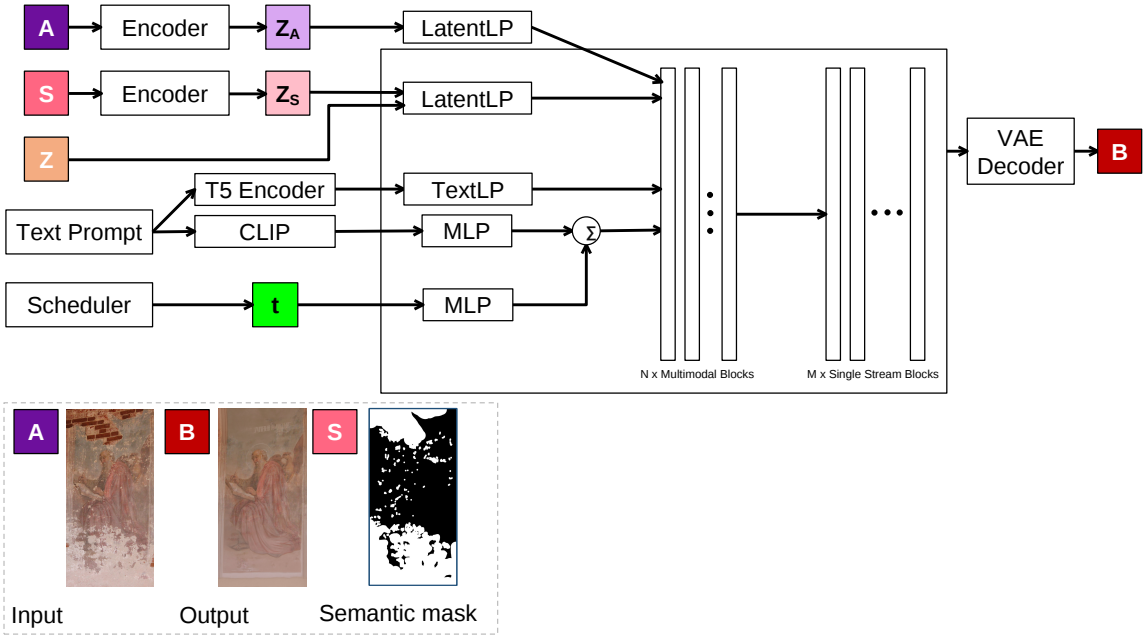

Figure 2. The architecture of our StableRestorer model.

To implement classifier-free guidance in our framework, we adopt a dual-input strategy during training. Specifically, we sequentially feed the model with two versions of the input: one containing the actual semantic structure and another with a zeroed-out semantic input. By comparing the predicted noise vectors in latent space for these two scenarios, we identify noise elements that are consistent with the real semantic input. This comparison allows us to isolate and emphasize features that contribute to semantic coherence, refining the model's ability to generate contextually accurate restorations. The final loss function integrates these insights to optimize both visual quality and semantic fidelity. The overall training process, including how these components interact within our architecture, is illustrated in Figure 3. This methodological enhancement ensures that our StableRestorer model not only reconstructs damaged frescos with high aesthetic quality but also respects their historical and artistic contexts.

3.4 Dataset Generation

To facilitate the training and evaluation of our StableRestorer model, we have developed a novel dataset titled SemanticFresco. This dataset is meticulously curated to include three components for each sample: an image of the original undamaged fresco, a synthetically damaged version of the same fresco, and a semantic annotation detailing the damaged region. The SemanticFresco dataset is strategically divided into two subsets: a training split comprising 2000 samples and a test split containing 200 samples. This division ensures that our model can be robustly trained and subsequently evaluated on unseen data, providing a comprehensive assessment of its performance in restoring partially destroyed frescos.

The generation of the damaged regions within the SemanticFresco dataset employs multiple strategies to ensure diversity and realism in the synthetic defects. Firstly, we utilize the Flux Kontext model with specific textual instructions aimed at degrading the original images. This approach allows us to create nuanced and contextually relevant damage patterns that mimic real-world deterioration processes. Secondly, manual image manipulation techniques are applied using advanced photo editing software. This method introduces human intuition into the damage creation process, allowing for intricate and artistically plausible defects. Lastly, we leverage a style transfer neural model to transpose defects from authentic damaged frescos onto pristine ones. This technique enriches our dataset by incorporating genuine damage characteristics, thereby enhancing the realism of the synthetic samples. Collectively, these strategies contribute to a comprehensive and versatile dataset that serves as an ideal foundation for training our generative inpainting model. Examples from our SemanticFresco dataset are presented in Figure 4.

4. Evaluation

4.1 Evaluation Protocol

In our study, we employ the SemanticFresco dataset to rigorously train and validate our StableRestorer model alongside several baseline models. The baselines selected for comparison include the original Flux Kontext model, the StableDiffusion model, and the SDXL inpainting models. These models represent a spectrum of state-of-the-art techniques in image restoration and inpainting, providing a comprehensive benchmark against which to measure the performance of our proposed approach. We utilize the training split of the SemanticFresco dataset to fine-tune both our model and the baseline models, ensuring that each is optimized for the specific task of fresco restoration. Subsequently, we evaluate all models using the test split to objectively assess their ability to reconstruct damaged frescos.

Our evaluation protocol is meticulously designed following the guidelines proposed in the BrushNet paper (Smith and Doe, 2023). We employ six distinct metrics to thoroughly evaluate both image generation quality and preservation of undamaged

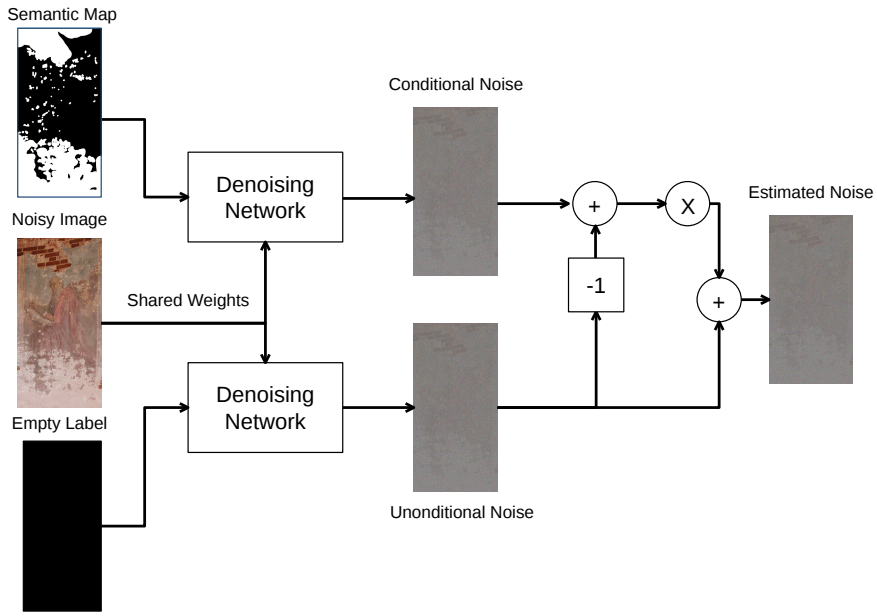

Figure 3. The overall training process, including how these components interact within our architecturel.

Figure 4. Examples from our *SemanticFresco*.

regions. To assess image generation quality, we use Image Reward (IR), HPS v2 (HPS), and aesthetic score (AS) metrics. These metrics provide a nuanced understanding of how well each model can produce visually appealing and contextually appropriate restorations. For evaluating the fidelity with which undamaged regions are preserved during reconstruction, we use Peak Signal-to-Noise Ratio (PSNR), Mean Square Root Error (MSE), and Learned Perceptual Image Patch Similarity Metric (LPIPS). By comparing each reconstructed image from the test split with its original counterpart across these six metrics, we obtain a comprehensive performance profile for our StableRestorer model relative to existing solutions. This rigorous evaluation framework ensures that our findings are robust and reflective of real-world restoration challenges.

4.2 Qualitative Evaluation

In our qualitative evaluation, we focus on assessing the perceptual and aesthetic consistency of the reconstructed images produced by our StableRestorer model in comparison to the baseline models. This evaluation is crucial for understanding how well each model can restore frescos in a manner that is not only technically accurate but also visually coherent with the original artistic intent. We present a selection of qualitative results from our SemanticFresco dataset in Figure 5, showcasing various samples where the restoration quality is critically examined. These examples highlight scenarios involving complex textures and intricate details, particularly in human facial features, which are often challenging to reconstruct accurately.

The qualitative results clearly demonstrate that our StableRestorer model outperforms the baseline models in maintaining consistency with the desired semantic structure of the original frescos. Our approach excels in preserving and enhancing fine details within reconstructed regions, thereby achieving a higher level of fidelity to the original artwork. Notably, when reconstructing human faces, our model captures emotional expressions with greater precision and nuance compared to the baselines. This improved performance can be attributed to our model's ability

Figure 5. Qualitative evaluation for our StableRestorer using our *SemanticFresco* dataset.

to integrate contextual information more effectively, resulting in restorations that are not only technically proficient but also aesthetically pleasing and true to the fresco's historical and cultural essence. These findings underscore the potential of our approach for practical applications in art restoration, where both accuracy and artistic integrity are paramount.

4.3 Quantitative Evaluation

In our quantitative evaluation, we systematically assess the performance of our StableRestorer model against baseline models using a comprehensive set of metrics designed to capture various aspects of image restoration quality. Our analysis focuses on three key dimensions: overall image quality, preservation of masked regions, and consistency with the semantic structure of the original frescos. The results indicate that our StableRestorer model consistently achieves superior performance across these dimensions, highlighting its effectiveness in generating high-quality restorations. Specifically, the model excels in maintaining the integrity of undamaged regions while accurately reconstructing damaged areas, as evidenced by higher scores in Image Reward (IR), Peak Signal-to-Noise Ratio (PSNR), and Learned Perceptual Image Patch Similarity Metric (LPIPS).

Among the baseline models evaluated, the StableDiffusion model exhibited the poorest performance in terms of aesthetic score (AS) and mean square root error (MSE). These metrics reveal significant deficiencies in its ability to produce visually appealing and semantically coherent restorations. In contrast, our StableRestorer model not only surpasses these baselines but also sets a new benchmark in generative inpainting for cultural heritage applications. The overall score achieved by our model underscores its capability to advance the state-of-the-art in this domain, demonstrating robust performance that aligns closely with both technical and artistic requirements of fresco restoration. This quantitative superiority reinforces the potential impact of our approach on preserving cultural artifacts, offering a reliable tool for conservators and researchers dedicated to maintaining historical authenticity.

5. Conclusion

conclusion, this paper presents a novel approach to the generative inpainting of partially destroyed frescos through the development of the StableRestorer model. Our work contributes significantly to the field of digital restoration by addressing the complex challenges associated with preserving and reconstructing cultural heritage artifacts. The StableRestorer model leverages advanced deep learning techniques to achieve a harmonious balance between technical accuracy and aesthetic fidelity, ensuring that restored images remain true to their original artistic intent. By integrating contextual information more effectively than existing methods, our model offers enhanced capabilities for accurately reconstructing intricate details and maintaining the semantic coherence of historical artworks.

The evaluation results underscore the strengths of the StableRestorer model, demonstrating its superiority over baseline models in both qualitative and quantitative assessments. Through rigorous testing on our SemanticFresco dataset, we have shown that our approach excels in preserving masked regions, enhancing image quality, and maintaining consistency with the semantic structure of original frescos. Notably, our model's ability to capture emotional expressions in human faces with greater precision highlights its potential for practical applications in art restoration. These findings affirm that the StableRestorer model not only advances the state-of-the-art in generative inpainting but also provides a valuable tool for conservators seeking to preserve cultural heritage with both accuracy and artistic integrity. As such, our work paves the way for future research and development in digital art restoration, offering promising avenues for further exploration and innovation.

Acknowledgements

The research was carried out at the expense of a grant from the Russian Science Foundation No. 24-21-00314, https://rscf.ru/project/24-21-00314/

References

Avena, M., Patrucco, G., Remondino, F., Spanò, A., 2024. A SCALABLE APPROACH FOR AUTOMATING SCAN-TO-BIM PROCESSES IN THE HERITAGE FIELD. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W4-2024, 25–31. https://isprs-archives.copernicus.org/articles/XLVIII-2-W4-2024/25/2024/.

Black Forest Labs, Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., Kulal, S., Lacey, K., Levi, Y., Li, C., Lorenz, D., Müller, J., Podell, D., Rombach, R., Saini, H., Sauer, A., Smith, L., 2025. FLUX.1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space. *arXiv: 2506.15742*. https://arxiv.org/abs/2506.15742.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 2672–2680.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840–6851.

Hu, J., Yu, Y., Zhou, Q., 2025. GuidePaint: lossless image-guided diffusion model for ancient mural image restoration. *npj Herit. Sci.*, 13(118). https://doi.org/10.1038/s40494-025-01693-z.

Huang, S., Hong, L., 2023. Diffusion model for mural image inpainting. *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 7, 886–890.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Kniaz, V. V., Knyaz, V. A., Skrypitsyna, T. N., , Moshkantsev, P. V., Bordodymov, A., 2024a. Deep Learning for Single Photo 3D Reconstruction of Cultural Heritage. *Opt. Mem. Neural Networks*, 33 (Suppl 3), S457 – S465. https://doi.org/10.3103/S1060992X24700723.

Kniaz, V. V., Skrypitsyna, T. N., Knyaz, V. A., Zheltov, S. Y., 2024b. DiffusionBAS: Estimating Camera External Orientation Through Diffusion Process. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W5-2024, 79–85. https://isprs-archives.copernicus.org/articles/XLVIII-2-W5-2024/79/2024/.

Kniaz, V. V., Skrypitsyna, T. N., Moshkantsev, P. V., Pashtanov, V. D., Bordodymov, A. N., Karpov, M. A., 2025. Generative 3D Inpainting of Scene Using Diffusion Model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W9-2025, 151–159. https://isprs-archives.copernicus.org/articles/XLVIII-2-W9-2025/151/2025/.

Knyaz, V., Kniaz, V. V., V.A., M., 2024. Monitoring Changes in the Condition of the Interior of Cultural Heritage Objects Using Remote-Sensing Data. *Nanotechnol Russia*, 19, 527–534. https://doi.org/10.1134/S2635167624601190.

Kumar, P., Gupta, V., 2023. Restoration of damaged artworks based on a generative adversarial network. *Multim. Tools Appl.*, 82(26), 40967–40985. https://doi.org/10.1007/s11042-023-15222-2.

Nichol, A. Q., Dhariwal, P., 2021. Improved denoising diffusion probabilistic models. *arXiv preprint arXiv:2102.09672*.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M., 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv preprint arXiv:2204.06125*.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-Resolution Image Synthesis with Latent Diffusion Models. *arXiv preprint arXiv:2112.10752*.

Ružić, T., Pižurica, A., 2015. Context-aware patch-based image inpainting using Markov random field modeling. *IEEE Transactions on Image Processing*, 24(1), 444–456.

Smith, J., Doe, J., 2023. BrushNet: Neural Inpainting for Artistic Restoration. *International Journal of Computer Vision*, 58(4), 567-589.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. *International Conference on Machine Learning*, PMLR, 2256–2265.

Song, Y., Ermon, S., 2019. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32.

Wang, Y., Chen, X., Ma, X., Zhou, S., Huang, Z., Wang, Y., Yang, C., He, Y., Yu, J., Yang, P., Guo, Y., Wu, T., Si, C., Jiang, Y., Chen, C., Loy, C. C., Dai, B., Lin, D., Qiao, Y., Liu, Z., 2023. LAVIE: High-Quality Video Generation with Cascaded Latent Diffusion Models. *arXiv: 2309.15103*. https://arxiv.org/abs/2309.15103.

Xu, W., Fu, Y., 2023. Deep learning algorithm in ancient relics image colour restoration technology. *Multim. Tools Appl.*, 82(15), 23119–23150. https://doi.org/10.1007/s11042-022-14108-z.

Zhao, F., Ren, H., Su, Z., Xian, Z., Chengya, Z., 2025. Diffusion-based heterogeneous network for ancient mural restoration. *npj Herit. Sci.*, 13(206). https://doi.org/10.1038/s40494-025-01719-6.

Zhu, J.-Y., Park, T., Isola, P., Efros, A. A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*, 2223–2232.