CoRe-GAN: A Gated Attention Framework for High-Fidelity Restoration of the Codex Borgia

MSc Research Project
Master of Science in Artificial Intelligence

Jose Manuel Lopez Garcia
Student ID: 24177075

School of Computing
National College of Ireland

Supervisor: Dr. Mohit Garg

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Jose Manuel Lopez Garcia
Student ID:	24177075
Programme:	Master of Science in Artificial Intelligence
Year:	2025
Module:	MSc Research Project
Supervisor:	Dr. Mohit Garg
Submission Due Date:	11/12/2025
Project Title:	CoRe-GAN: A Gated Attention Framework for High-Fidelity Restoration of the Codex Borgia
Word Count:	8782
Page Count:	25

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	25th January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

CoRe-GAN: A Gated Attention Framework for High-Fidelity Restoration of the Codex Borgia

Jose Manuel Lopez Garcia
24177075

Abstract

This project presents CoRe-GAN, a gated-attention generative model for restoring damaged regions of the Codex Borgia, a highly detailed Mesoamerican manuscript. A custom dataset of segmented codex imagery was created, along with a CodexDamageMaskGenerator that simulates realistic deterioration such as cracks, pigment loss, and abrasion. CoRe-GAN combines gated convolutions, contextual attention, and multi-stage training to address the manuscript's complex iconography. Quantitative and qualitative evaluations show that the model significantly outperforms baseline inpainting methods in PSNR, SSIM, LPIPS, and texture fidelity. The results demonstrate the value of domain-adapted AI for cultural-heritage restoration.

1 Introduction

Preserving the cultural heritage by restoring art pieces, murals, paintings and all kind of iconography is vital to keep humankind historical legacy. These artistic expressions can suffer damage from diverse causes, endangering its safeguarding (Kucuk et al.; 2024).

Among these cultural heritage artifacts, emerges the Codex Borgia, which is one of the most sophisticated and complex artistic expression to survive from Mesoamerica. Mesoamerica refers to the prehispanic cultural area of the current territory of Mexico, and some parts of Honduras, Belice, Guatemala, El Salvador, Nicaragua and Costa Rica. The Codex Borgia is heavy in symbolic value, its pages contain depictions of deities, rituals, calendrical structures, cosmological cycles and glyphs. The Codex Borgia is one of the most important sources for anthropologists and archaeologists alike to have a little insight into the civilization that created it, is a window to the past. The document originates from the late Postclassic period and its manufactured from deerskin and paper amate. Its antique nature along with the characteristic of the materials used in its manufacture, have caused great deterioration to the document. Among the damage there is abrasion, flaking, cracking and color fading. The vastly used methods for restoration are made by hand by artists, these processes can be time consuming and not always satisfactory (Buvaneshwaran et al.; 2025). Some experimental studies are now using generative AI regarding artwork restoration to improve the speed, precision and, scalability upon the traditional methods (Ansari; 2025).

Recent research in deep learning, particularly for image inpainting using Context Encoders, Convolutional Neural Networks(CNNs) and Generative Adversarial Networks(GANs), have shown that there is a great potential for automated restoration tasks. Context Encoders performs predictions that matches their surrounding areas, while GANs aim to improve both color and texture of the image(Buvaneshwaran et al.; 2025). Other approaches, such as Conditional GANs like Pix2Pix have been investigated, however the outputs from these models can possess stylistic differences from the original, which can compromise the historical accuracy of the restoration(Ansari; 2025). Quantitative analysis have been made to compare the AI restorations with the ones produced manually. Some of these metrics that compare the image quality are Peak Signal to Noise Ratio and Structural Similarity Index(Leong; 2025a).

Despite the progress in deep learning inpainting restoration, the majority of the existing models are trained with datasets that contain modern photography or synthetic damage that does not coincide with the one found in cultural heritage documents. More over, when there has been researches that tackle the issue of restoring artistic expressions the datasets that have been used have not enough representation from non-Western art. This lack of diversity can cause cultural bias (Ansari; 2025). The resulting cultural bias of Western developed AI models can reinforce cultural inequalities. That is why the use of more diverse datasets can mitigate this issue(Leong; 2025b). It also worth nothing that there has not been an attempt to apply AI models to restore Mesoamerican codices.

To address these limitations, this project introduces CoRe-GAN, a domain adapted restoration framework that is specially designed to deal with reconstruction of Mesoamerican pictorial manuscripts. The model combines three innovations that are in sync with the characteristics of the Codex Borgia. First, in uses a gated U-Net architecture with a contextual attention, this allows the system to selectively pass relevant features while it borrows non local textures from regions that do not suffer from damage. This is specially useful when reconstructing repetitive motifs, geometric borders and complex iconography. Second, it introduces a Codex-specific damage simulator, which create masks that simulate the real deterioration of the codices. These allows the model to learn from training conditions that emulate cultural heritage degradation, rather than synthetic environments. Lastly, it follows a multi stage training curriculum that is designed with the aim of gradually guiding the generator from the structural reconstruction to fine grained texture synthesis.

By customizing the model architecture and the training conditions to the features present in Mesoamerican codices, this project aims to not only improve a quantitative performance, but also to explore how can a machine learning model can engage with the visual logic of cultural heritage sources. The resulting model was evaluated and compared against baseline models, and its performance showed achieve state of the art restoration quality across both quantitative and visual analysis. More over, this works proposes a methodological approach that can be used for developing restoration models that are sensitive to the cultural significance of the artifacts, contributing to inpainting image research as well to the preservation of the pictorial heritage of ancient civilizations.

In addition to this, the project created a custom made dataset from the Codex Borgia, first of its kind for AI restoration. By using a dataset consisting of Mesoamerican art and adjusting the AI model to better predict this particular art style, the research will contribute not only to the deep learning Art and cultural heritage restoration models by mitigating cultural bias, but also to the preservation of the cultural heritage of the area. The information retrieved from this research can help for the further development of projects that amalgamate historical, art, and archaeological research with the generative AI technologies. In addition, the creation of the dataset, will help future researchers to have an already curated data repository. This will ease future research.

Research Questions: How can generative AI—specifically gated-convolutional GAN architectures—be adapted to restore damaged regions of the Codex Borgia while maintaining stylistic, chromatic, and structural authenticity? Which computational approach among contemporary inpainting methods (Context Encoders, CNN-based models, GAN-based architectures) provides the most accurate and historically faithful restoration results for Mesoamerican manuscript imagery?

Research Objectives: Construct a curated, high-quality dataset of segmented regions from the Codex Borgia suitable for GAN-based training. Design and implement CoRe-GAN, a restoration model informed by the structural complexity and stylistic features of Mesoamerican manuscripts.Develop codex-specific synthetic deterioration masks that realistically simulate damage patterns.Evaluate CoRe-GAN against baseline deep learning models using both pixel-level and perceptual metrics. Contribute methodological insights and datasets that support future AI-driven restoration research in non-Western art traditions.

The structure of this report is organized as follows: Section 2 reviews prior work in classical, deep learning, and cultural heritage inpainting methodologies. Section 3 presents the methodology. Section 4 details the architecture design. Section 5 details the implementation. Section 6 reports quantitative and qualitative results and provides an in-depth evaluation and discussion. Section 7 concludes the study and proposes directions for future research.

2 Related Work

There have been various researches in restoration of art pieces with the use of AI models. As mentioned before, the vast majority of all the projects work with datasets of corresponding to western art. These project proposal will address the most relevant works for the investigation.

2.1 Classical Image Inpainting

The techniques used in early image inpainting were based on principles from partial differential equations (PDEs) and texture synthesis. Some of the approaches were Diffusion based like the ones made by Bertalmio et al. (Bertalmio et al.; 2000) and Telea (Telea; 2004). Their researches solved PDEs over the damaged section of the image, achieving the propagation of local intensity information into the missing regions of the image. These methods were efficient in computation and effective for small and narrow defects, however there was a tendency to over smooth textures and a failure to reproduce complex structural details.

Other approaches like the Exemplar based methods, such as the one proposed by Criminisi et al. (Criminisi et al.; 2004) addressed these shortcomings by coping texture patches from non damaged ares into the ones that suffered from it. This proved to be successful for structured textures and repetitive patterns, however, these patch methods underperformed when there are not suitable exemplars or when the missing regions contained semantically meaningful components that are unable to be inferred from a local context. These limitations make the classical techniques for inpainting not suitable for our project objective, due to the incapability of restoring complex iconography like the one present in the Codex Borgia. Mesoamerican manuscripts have a complex structure, with fine line works and pigment variations.

2.2 Deep Learning for Image Inpainting

As convolutional neural networks (CNNs) were developed, so it did the inpainting restoration techniques. They enabled the models to learn semantic representation instead of depending only on the propagation of local textures. The research made by Pathak et al. (Pathak et al.; 2016) was one of the pioneers by proposing a Context Encoder framework, which enclosed inpainting as an encoder-decoder reconstruction task being supervised by pixel losses. This approach yielded results capable of predicting a probable global structure, but frequently produced blurry textures caused by the limitations of the losses used (L1/L2) when modeling details of a high frequency nature.

On the other hand, Generative Adversarial Networks(GANs) made a great advance realism in inpainting by adding an adversarial objective, which promoted better and sharper textures while keeping a stylistic coherence. Izuka et al. (Iizuka et al.; 2017) proposed an inpainting model that was consistent locally and globally, with the use of dual discriminators, improving the fidelity of textures and the semantic reasoning. Another distinguished research was done by Yu et al. (Yu et al.; 2018) with the introduction of contextual attention. This approach enabled the networks to borrow and rearrange textures that were relevant from surrounding areas in a comprehensive manner. This type of methods are highly valuable for artwork restoration as detailed patterns and repetitive motifs are common.

Therefore, GAN based approaches have been significantly used for artwork restoration due to being able to learn mappings between a damaged image and clean one, with the use of paired or unpaired datasets. Buvaneshwaran et al. (Buvaneshwaran et al.; 2025) developed a restoration method which combines a Context Encoder with a GAN refinement phase. The pipeline followed the preprocessing of damaged images, predicting of missing content by using a convolutional autoencoder and the refining of the outputs with the help of a ground truth consisting of a GAN trained in historical datasets. Their generator is a standard encoder-decoder design with ReLU activations, meanwhile, the discriminator operates as a binary classifier with the help of convolutional layers. They achieved good results, a high Peak Signal to Noise Ratio (PSNR) but their training relied only in synthetically masked data and their system underperformed in generalization across different art styles.

Some other projects, like the one developed by Li et al., prefer to use a Conditional GAN (cGAN) as it has some advantages over a plain GAN (Li et al.; 2018). GAN can be unstable during the training process. This results in the output having some artifacts like noise or color shifting in the generated images. When there is an incorporation of conditional information in the GAN, it results in a more effective learning process. The conditioning variables of the model increases the stability of the learning process and thus improves the generator capability of representation. The cGAN varies from the GAN algorithm because cGAN learns to generate an clear image from an input and random noise by optimizing its objective function. This approach has obtained great results in image fields such as image inpainting, style transfer and super-resolution (Li et al.; 2018).

As we can observe, these works demonstrate that GANs have a good capacity to model the semantic structure and secure a stylistic consistency, both cornerstone properties for cultural-heritage restorations.

2.3 Free-Form Mask Inpainting and Gated Convolution

One of the main limitations of GAN based models for restoring images was their need for regular and axis aligned masks. However, the work done by Yu et al. (Yu et al.; 2019) tackled this issue by incorporating free-form masks and gated convolutions, which work with a dynamic modulation of the feature propagation based on the validity of the masks. These gating mechanisms provide the ability for the model to learn mask aware features while also handling arbitrary damage like shapes. This approach is suitable for restoring real deterioration like the one found in ancient manuscripts, such as codices. In addition to the later mentioned, the use of spectral normalized PatchGAN discriminator as well as multi objective losses, stabiliseed the training and also improved the realism of the textures generated. Gated convolutions have gained territory when doing modern inpainting frameworks, is thus, is the direct inspiration for the architecture that we implemented in this project.

2.4 Damage Detection and Restoration with CNNs

Other than their use in generative models, CNN-based pipelines have been used to detect, classify, and restore damage in cultural artifacts. Their main use is to detect where the image needs to be intervened. Also CNN has been used for assessing the aesthetic quality of artwork (Ansari; 2025).

Yan et. al used internet of things to detect the conditions of murals and obtain data to pass to their model. They proposed a model termed as ArtDiff, which enhances the U-Net architecture and introduced what they call FCST-NET for the localization of the cracks and blank spaces in the murals. This model uses the advantages of the Swin Transformer to improve the feature extraction and representation capabilities. The model then introduces a CE convolutional layer to compress the resolution of the input to one sixteenth of its original size. That way they can extract a more compact feature representation. For the encoder their proposal uses a Swin Transformer which processes the image in layers. Then the compressed feature maps are passed to the decoder which uses a deconvolutional upsampling model. They also used a deep supervision strategy which involves additional supervision signals at various levels, according to the authors, this helps the network to better learn the features, improve the training and accelerate the model convergence. The multilevel supervision has the aim for the model to learn about the cracks features during all the stages, enhancing their detection and reparation.More over, they used edge guided strategy. However this method that used retained edge details to repair the cracks, may produce noticeable deformations in the fillings. Finally, they perform the digital restoration of the art by using a diffusion based model(Yan et al.; 2025).

The project does not clearly specify what dataset was used, it only mentions that it was murals from diverse historical eras, this limits its replicability.

2.5 Diffusion Models for Artwork Restoration

On the other hand, some authors mention that the use of CNN inpainting models have its limitations. It is mentioned that an efficient CNN model needs a large amount of data to be trained. The problem that this represents is originated from the different unique art styles, which can be lost when a CNN is used for restoration. In addition, the damage of the art pieces, manifest as noise and outliers in the images. This may affect the performance of the CNN. An advantage is that diffusion models have a great effectiveness in handling and reducing noise and outliers in the images. Diffusion models also conserve the artistic style because the restoration that it performs is based on the local characteristics of the image (Huang and Hong; 2023).

The process of a diffusion model includes two phases. The first is known as forward diffusion. It occurs when the original image is modified with progressively added noise to become more blurry. Then amount of noise added is known, which allows for recursive calculations. The core of this stage is the relation with the noise. Then the second stage is known as reverse diffusion. It starts with the noisy image which gradually goes back to the original clean state (Xu et al.; 2024).

Usually, Diffusion models require large datasets to achieve good results, however the authors trained a diffusion model using a collected dataset of 1000 murals, but they compensated this by training for 20,000 times. Even so, the results were not as good as expected, as the model performed better with the celeba dataset, a large one, as it was pre trained on it. They proceeded to fine tune the model with a Guided Diffusion and their metrics improved. They obtained a PSNR of 22.31 (Huang and Hong; 2023). Diffusion models are one of the modern alternatives for inpainting as gives excellent results at texture preserving restoration, however, they require high computational cost and a big sample, both big limitation factors for our project.

2.6 Deep Learning for Cultural Heritage Restoration

The source of the dataset used in this project, the codex Borgia, is a document of high cultural and heritage value. Works that apply deep learning to heritage restoration are growing among the research community, showing promising results when applied to frescoes, murals and paintings. There is a variety of approaches to this matter, among them, we can find the works done by Liu et al. (Liu et al.; 2024) applied a frequency and multi layer feature enhancement networks for ancient painting restorations. A different solution was made by leveraging coordinated attention for detailed mural restoration, named CAAT-GAN by Zhang et al. (Zhang et al.; 2025). Merizzi proposed a unsupervised Deep Image Prior (DIP) framework for the reparation of frescoes which remain texture consistent (Merizzi et al.; 2024). A two stage line drawing guided approach was taken for reconstructing structural outlines was explored by Li et al. (Li et al.; 2024). Zhao et al. (Zhao et al.; 2024) proposed a GAN - transformed hybrid model for heterogeneous mural restoration.

As we can observe, all these works have a different approach to the restoration of cultural heritage imagery, and are a meaningful progress in the field. However, the majority focus on murals, which generally possess large scale, homogeneous textures. On the other hand, manuscripts such as the Codex Borgia, contain in its pages fine details, symbolic motifs and saturated pigments, which require a model that adapts to these characteristics and is capable of both a precise structural reconstruction and stylistic coherence.

After going through the related work on the field, we can observe that the literature demonstrates that there is a constant advance in the field of image inpainting restoration, however there are still persisting methodological limitations. For an instance, classical approaches fail to reconstruct complicated semantic structures, and early CNNs methods result in blurry texture due to their reliance on pixelwise loses. With the introduction of GAN models there was an improvement on perceptual realism, but they often depend on synthetic masks and show an instability while training, resulting in a limited generalization when dealing to real cultural heritage damage. Meanwhile, Diffusion models can tackle some of these issues, however they represent a high computational cost and depend heavily on pretrained natural image distributions, which are considerably different from historical artwork. More over, the cultural heritage studies have their primary focus on murals, which neglects other cultural expressions such as Mesoamerican manuscripts, which haver their own structural an stylistic complexity. Adding to this, a lack of publicly available datasets and domain-specific deterioration models restricts even more the reproducibility and applicability of these experiments. The present project addresses these gaps with a focus on a gated convolutional inpainting method, specifically tailored to the Codex Borgia contents, and it is supported by a custom dataset and a realistic image simulation framework.

3 Methodology

The goal of the project is to demonstrate the feasibility of applying modern generative models to aid in the preservation of Mesoamerican codices. This area of study is largely unexplored as seen in the previous literature review section. Despite the extensive work that has been made in painting and mural restoration, there is still not a comparable deep learning research for Mesoamerican manuscripts, like the Codex Borgia. These documents present unique challenges due to their composition, its symbolic iconography is highly intricate, the motifs are dense and repetitive, the line work tends to be thin and detailed and their mineral pigments are saturated. In addition to their innate characteristics, the damage they present is non uniform. Existing methods lack to attend this issues because they typically assume mural like textures or rely on synthetic training data that not coherent with manuscript specific deterioration.

To address these gaps, the present project takes the following measures.

- Constructing the first dedicated Codex Borgia dataset
- Implementing a gated convolution + contextual attention inpainting model adapted to the characteristics of the document
- Designing a Codex-specific free form damage simulation
- Using a multi stage GAN with LPIPS Losses strategy for adversarial and perceptual training tailored to cultural heritage images to achieve both semantic and stylistic accuracy.

Therefore the proposed restoration framework, that we decided to name as CoRe-GAN (Codex Restoration Network with Gated Attention) was developed as domain adapted inpainting architecture specifically designed to handle the restoration of Mesoamerican codices. This positions the project as a novel contribution not only to computer vision but also to the digital preservation of Indigenous cultural artifacts.

It is worth mentioning that before settling on the final CoRe-GAN architecture, the project evolved through a number of exploratory stages that aided in the understanding of how different inpainting strategies behave on the Mesoamerican dataset. Two baseline models were implemented, a Context Encoder and a UNet-PatchGAN. Their selection was not arbitrary, because both of the models represent two widely used approaches for in painting. The Context Encoder is a classical encoder decoder baseline that can handle rough structural completion. On the other hand, UNet-PatchGAN goes further by incorporating skip connections as well as adversarial supervision, characteristics that are considered as the foundations in modern restoration pipelines. By including these models as baselines, it allowed the project to have reference points that reflect increasing architectural complexity and helped to identify which components were necessary to achieve a successful restoration of the images.

Based on their limitations, primarily reduced texture fidelity and unstable GAN convergence, the final model to be described in this chapter was selected. A detailed comparison of these experimental variants is presented in the evaluation section. However, the study acknowledges the important role of these architectures as stepping stones to develop CoRe-GAN.

3.1 Dataset

The dataset used for this project had to be created. This decision was made because currently, there is not an available one that fulfills the requirements needed for a high resolution restoration. The original data this project were obtained from two different repositories. One was the Mediateca of the National Institute of Anthropology and History of Mexico (INAH by its name in Spanish). The INAH resource contains multiple documents that were digitalized with the aim of promoting research. The second source of data was be collection of the FAMSI website.

The custom data set for the project was created using the Codex Borgia, which is a 76 pages long document from Mesoamerica. This codex contains ritual scenes of the antique religion, depictions of the most important deities and a calendar section. The codex originates from the now territory of Mexico. However, this manuscript has been victim of damage caused by time and other factors, therefore its actual condition suffers from pigment deterioration, fissures, abrasion marks and color fading. To overcome this situation, the codex was carefully inspected, and each page was manually segmented into smaller sub-regions that posses a clean, clear and meaningful visual importance. This process yielded 968 images for model training. The approach of of creating a custom dataset is a regular practice when dealing with a restoration research were cultural-heritage content is the main asset, these custom datasets are required because the deterioration patterns of these historical documents or artifacts are unique (Zhao et al.; 2024; Liu et al.; 2024).

In addition to this training dataset, 25 images were reserved with the goal of evaluating the model behavior when prompted with unseen data.

3.2 Image Preprocessing

The images used in the project had to be preprocessed to keep its consistency throughout the training. They were cropped randomly in order to remove borders, margins and enrich the dataset. They were all resized to 256 x 256 resolution, which is consistent with the fixed input size requirement of the Gated U-Net generator used in the project. This resolution allowed the model to have an efficient training and at the same time, preserving enough detail for the reconstruction of the iconography. The images were transformed with the previous mentioned resize, the conversion to tensor and a normalization to [-1,1].

$$x_{\text{norm}} = 0.5x - 0.5$$

The standardization of the images supports a stable GAN training and keeps its compatibility with both perceptual and LPIPS networks. The process of normalization and keeping a consistent size is a well established step used in preprocessing pipeles for GAN based inpainting for image restoration (Yu et al.; 2018, 2019).

3.3 Synthetic Damage Mask Generation

Synthetic masks were applied to the images to simulate deterioration. The project used two mask generators, one creates irregular Free-form masks, and one Codex-specific damage masks. The training process used them in different instances of the process with the aim of achieving better results.

3.3.1 Irregular Free-Form Masks

This generator creates masks that are formed of brush strokes. They form a random free form built from 5 to 15 strokes with 4 to 12 vertices per stroke. The masks define the missing region to 0 whereas the visible or known region to 1. This type of masks were used for the initial model learning, and simulate generic irregular damage patterns. This approach is similar to the one used in recent free-form inpainting works (Yu et al.; 2019).

3.3.2 Codex-Specific Damage Masks

Once the initial training was complete, the model went into a fine tuning phase. For this, a mask generator named CodexDamageMaskGenerator which generates more realistic masks that simulates the damage found in old manuscripts was used. The generator used the resources of blobs to simulate pigment loss or flaking, cracks for fissures that usually follow a trajectory, abrasion strokes for scrapes similar to the one caused by a brush, fading clouds for eroded areas; and mixed masks to emulate complex damage situations. The masks were only used when they represented a pixel coverage between the range of 3-35% of the total image, this with the purpose of ensuring a degradation that is severe and concise with real damage. The masks were generated dynamically for each mini-batch. The project took inspiration for this technique from similar researches where a design of domain specific synthetic damage was used to train the restoration networks, and it obtained positive results in restoring ancient murals and frescoes (Li et al.; 2024; Zhang et al.; 2025; Zhao et al.; 2024).

3.4 Generator Architecture: Gated U-Net with Contextual Attention

In this project it was decided to implement a generator that is a Gated U-Net with a fixed resolution. The generator main components are a 6 layer encoder that uses a Gated Convolution which takes advantage of learnable soft masks that controls the information flow, a bottle neck Contextual Attention block that enables the non local features borrowing from regions that does not present damage, a 5 layered decoder that uses Gated Deconvolution along with skip connections and a final 3 x 3 convolution + Tahn activation in order to produce an [1,-1] output. These elements follow previously implemented architectures used for inpainting research.

The Gated Convolution was chosen for the model because, instead of a regular convolution, each layer learns gating function that is feature wise that can be defined as:

$$\text{Output} = \text{FeatureConv}(x) \odot \sigma(\text{GateConv}(x))$$

This allows the network to control which features need to be passed forward, adapting it along the way. FeatureConv(x) produces the primary representation of the features, while σ(GateConv(x)) is in charge of generating gates that are element wise that decrease or increases these features selectively. The model achieves a better flexibility when using this gating structure while also being able to effectively suppress information that is deemed irrelevant. This translates to a better capacity of capturing patterns that are relevant for the task. The generator is based of the Gated Convolution architecture that was introduced by Yu et at. (Yu et al.; 2019). This approach, as mentioned before, replaces standard convolution with a gating mechanism that learns from itself, and this is reflected in a better modulation of the features propagation depending on the mask validity. This translates to the network being mask aware, a technique that has yielded good results for free form inpainting restoration performance.

Using the Contextual attention block allows the model to implement a module that deals with non local attention, using features from unmasked regions to reconstruct the textures that are missing. This approach is highly important when geometric patterns are present in the codex's iconography, because it captures the long range dependencies and the patterns that repeats constantly. The module takes inspiration from the research made by Yu et al (Yu et al.; 2018).

Moreover, the use of Skip Connections ensures that the encoder features are combined with the features of the decoder to successfully restore spatial details, maintaining high resolution details.

3.5 Discriminator: Spectrally Normalized PatchGAN

The model uses a PatchGAN discriminator that classifies patches of size 70 x 70 that are overlapping as real or fake. The PatchGAN formulation follows the work done by Isola et al. (Isola et al.; 2017), which outputs a grid of real and fake decisions over local image patches. PatchGANS are used widely in inpainting tasks because they promote high frequency realism and also sharp textures.

The use of Spectral normalization on the discriminator stabilizes the training by forcing the Lipschitz constant of each of the convolution layers. The discriminator then outputs a spatial grid of logits instead of a singular fake or real score. This enables the adversarial supervision to be fine grained, improving the realism of the texture. This technique of normalization follows the ones used in Yu et al. (Yu et al.; 2019).

This particular configuration has been used with successful results in restoration studies applied to murals and frescoes (Zhang et al.; 2025; Merizzi et al.; 2024).

3.6 Loss Functions

The generator of the project was trained following the structure of recent researches of inpainting methods that use a multi term loss that combines reconstruction, perceptual and adversarial terms (Yu et al.; 2019; Zhang et al.; 2018). The training objective combines four loss components:

- **Reconstruction Loss (L1):** $\mathcal{L}_{\text{recon}} = \|M \odot (G(x) - y)\|_1$ (masked pixels only).

- **Perceptual Loss (VGG-19):** $\mathcal{L}_{\text{perc}} = \|\phi(G(x)) - \phi(y)\|_1$ using VGG-19 up to `relu4_1` (Johnson et al.; 2016).
- **LPIPS Loss:** $\mathcal{L}_{\text{LPIPS}}$ following (Zhang et al.; 2018).
- **Adversarial Loss (PatchGAN):** $\mathcal{L}_{\text{adv}} = \text{BCE}(D(G(x)), 1)$.

Final generator loss: $\mathcal{L}_G = \lambda_r \mathcal{L}_{\text{recon}} + \lambda_p \mathcal{L}_{\text{perc}} + \lambda_\ell \mathcal{L}_{\text{LPIPS}} + \lambda_a \mathcal{L}_{\text{adv}}$. The discriminator uses standard adversarial BCE loss. Loss weights: $\lambda_r = 100,\ \lambda_p = 10,\ \lambda_\ell = 10,\ \lambda_a = 1$ (with λ_a later reduced during fine-tuning).

3.7 Training Procedure

The training process of the model can be divided into three phases: The Context Encoder Pretraining, the full GAN Training; and the Fine Tuning Strategy. This follows best practices present in inpainting research (Yu et al.; 2018, 2019).

3.7.1 Context Encoder Pretraining

During this first phase the generator was trained without the discriminator. This stabilizes training by allowing the network to learn rough structural reconstruction using only L1, perceptual, and LPIPS losses. The phase ran for 60 epochs; and trains the network to predict the missing regions without having to deal with the adversarial pressure. While this phase runs, the best model is tracked by the PSNR and it is saved automatically.

3.7.2 Full GAN Training

Once the rough predictions are stable, the adversarial training can be introduced. This phase corresponds to 400 epochs. Now, the discriminator was activated which adds the adversarial supervision. With each iteration the discriminator and the generator were updated and the corresponding metrics of generator loss, discriminator loss, PSNR, L1, Perceptual and LPIPS components were logged. The PatchGAN discriminator leads the generator to create more realistic textures and a sharper reconstruction. This mirrors methodologies used for modern art restoration (Zhang et al.; 2025; Li et al.; 2024). The best model was tracked and saved as well as automatic checkpoints and visual outputs were generated at regular intervals.

3.7.3 Fine Tuning Strategy

This phase consists of five stages. This approach was taken to improve the restoration of codex like damage. The first stage ran for 50 epochs, it reduced the learning rate to 5×10^{-5} and applies the full adversarial weight. The second stage corresponds to 50 epochs and it lowered the adversarial weight, which translated to a very stable training that focused on structural consistency. The third stage ran for 30 epochs and made emphasis in being more precise on edge and small detail restoration by doubling the L1 weight. The fourth stage aim was to replicate a subtle codex degradation and refine micro textures by using a softer mask generation. The final stage trained using only the CodexDamageMaskGenerator for 100 epochs. This approach helped the generator to learn degradation patterns unique to the codex. This approach that seeks the transition from rough structure refinement to fine textures synthesis by learning rate schedules and loss-weight adjustments to guide the generator, is similar to the ones used in heritage restoration works (Zhao et al.; 2024; Merizzi et al.; 2024).

During all the fine tuning, the best model was tracked and regular checkpoints and outputs were saved. There was a distinction when saving the best model from stage one to four and the final one. The best epoch was saved as best_G_finetune and best_G_codex100 respectively. This separation was decided for future comparison of the model performance with and without the Codex like mask training.

3.8 Evaluation Metrics

The metrics that were used are Peak Signal-to-Noise Ratio (PSNR) which evaluates how accurate the pixel reconstruction is. Learned Perceptual Image Patch Similarity (LPIPS) which measures the perceptual similarity. LPIPS is particularly relevant because recent heritage restoration literature has put more weight on perceptual quality over pure pixel accuracy(Zhang et al.; 2018; Zhao et al.; 2024). Mean Absolute Error(MAE) gives insight of pigment level deviation by measuring the average absolute error. L1 Reconstruction Error (L1) reflects how well a model minimize the reconstruction objective during training. Structural Similarity Index(SSIM) assesses the structural integrity, which is helpful to preserve iconographic patters. In addition to this, a visual inspection of the quality of the image was made. Grids that shows the original, masked and restored image were generated to aid with the visual analysis.

4 Design Specification

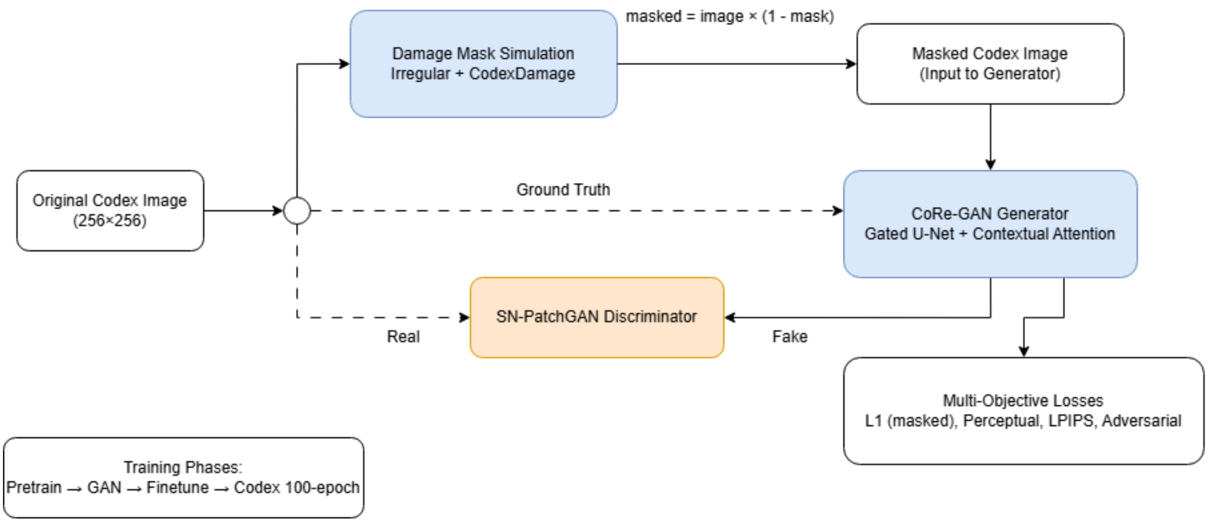

Figure 1: System Overview of the CoRe-GAN Framework

The proposed CoRe-GAN model combines gated convolutions, contextual attention bottleneck, and a U-Net–style encoder–decoder structure to effectively model both the structural organization and the intricate textures characteristic of the Codex Borgia. The generator operates as a mask-aware network in which the gating mechanism selectively modulates feature propagation according to the spatial validity of the input, while the contextual attention module enables non-local texture borrowing from undamaged regions. A PatchGAN discriminator with spectral normalization provides adversarial supervision to enforce high-frequency realism. The model is trained using a multi-stage curriculum consisting of context-encoder pretraining, full adversarial optimization, and two phases of domain-specific fine-tuning. Crucially, the network is trained not only with generic free-form masks but also with a custom CodexDamageMaskGenerator, which simulates pigment flaking, fissures, abrasion strokes, fading regions, and mixed complex deterioration patterns derived from observed degradation in historical manuscripts. This combination of architectural design and culturally informed damage modeling forms the technical foundation of CoRe-GAN.

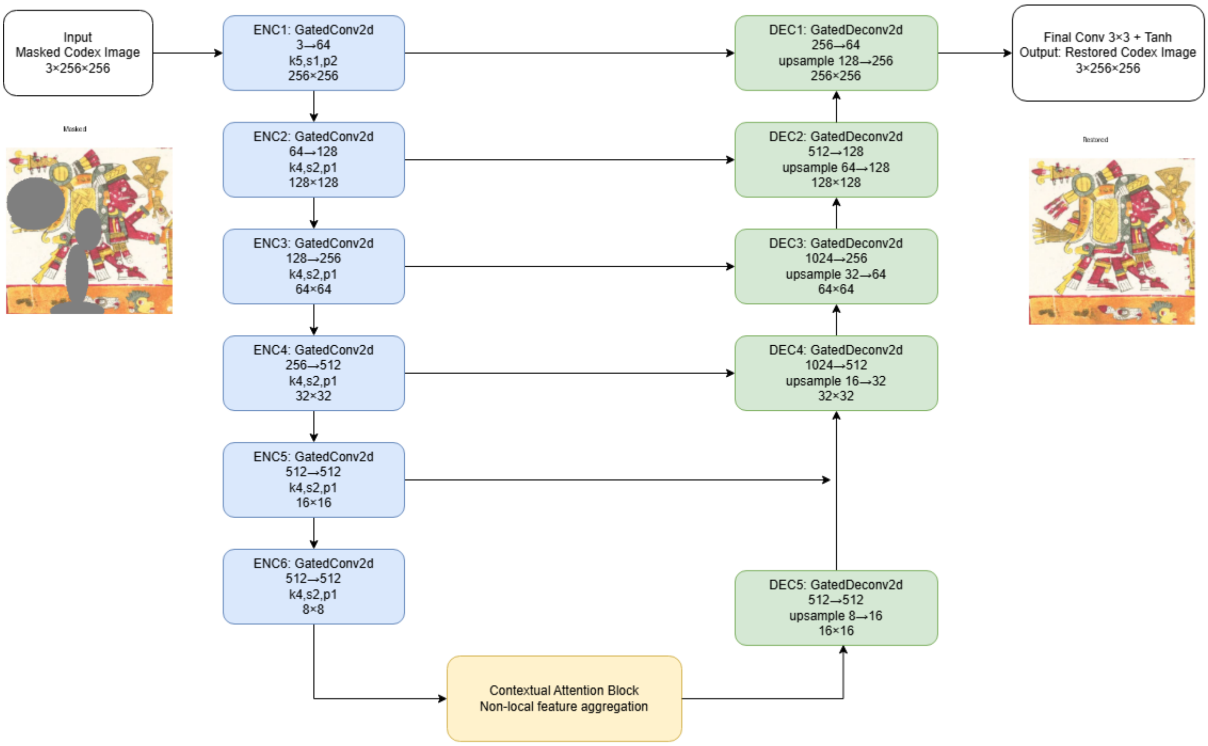

Figure 2: CoRe-GAN Generator Architecture

5 Implementation

The complete project, including model training, development and inference was built in Python 3.10 using PyTorch. All experiments were executed in Google Colab notebooks with NVIDIA L4 GPU acceleration.

The restoration models were implemented inside a unified PyTorch based inpainting network, which was specially designed to meet the requirements of a damaged Mesoamerican codex. These historical documents suffer from deterioration patterns like cracks, abrasions and flaking pigments, which make them different from the typical natural image datasets. To being able to simulate this degradation scenarios, a custom mask generation pipeline was developed and applied to all models. This was made with the objective to compare the performance between the architectures while also maintaining realistic conditions faced by these pre Columbian pictorial documents. There was a total of 25 images used for the inference module, representing unseen data.

The system resized all inputs to 256 x 256 pixels and were normalized to a range of [-1,1]. A single inference pipeline was used for all the architectures to keep full procedural uniformity. The pipeline loads the original version of every image, applies a mask, produces the corrupted input and feeds it to the corresponding generator. A standardized visualization grid (Fig. 3) is created showing: Original Image, Mask, Masked Input and restored Output.

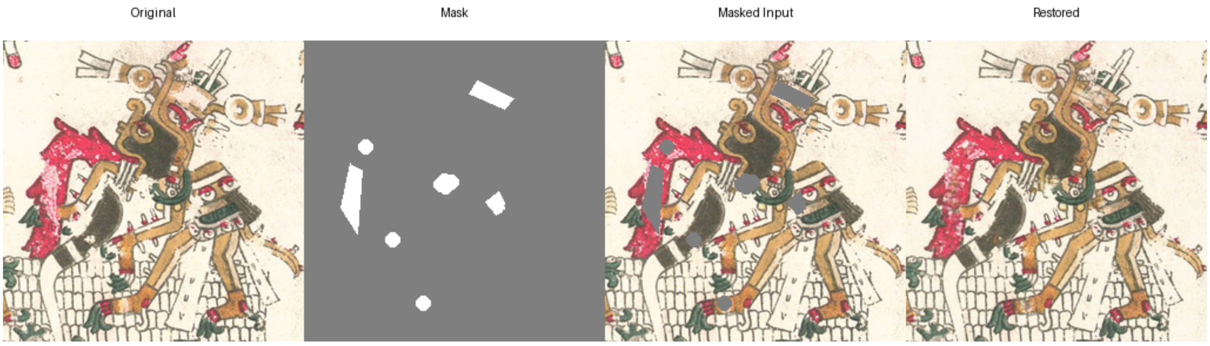

Figure 3: Visualization grid: Original, Mask, Masked Input, and Restored Output.

As mentioned before, every image was mated with a synthetically degraded version that was applied one of three mask types. The first type is the balanced stroke masks, which aim to emulate pigment loss caused by abrasion or accidental or intentional scrapping, common in ancient codices that suffer damage from diverse causes. This masks were created from a collection of irregular strokes with a randomized vertex counts, angles and thickness. These parameters yielded organic like shaped regions of missing pigments. The second type is the cracks masks which emulate fissures that usually form in aged organic materials, such as deer skin, which is of what the Codex Borgia is made of. The mask generator draws branching crack networks with randomized angles, lengths and local perturbations, such as the damage found in brittle aged surfaces. This type of damage is common on Mesoamerican codices which often present micro fractures. Finally, manual masks were drawn by hand, to reflect actual damage present in the codex. These masks are specially important as they give us a direct insight on how the models perform on real case scenarios faced by archaeology when restoring codices. All the masks (Fig. 4)were post processed using a morphological erosion scheme to limit the total coverage to approximately 18% of the image, this kept the degradation severity constant across different masks and models. This was designed to maintain a controlled experimental condition.

The quantitative evaluation was conducted in a same manner for all model using five metrics: PSNR, SSIM, LPIPS, MAE and L1 reconstruction loss. The metrics were recorded per image, per mask type and per model into CSV files that were later used for statistical and comparative analysis.

This unified approach allows to observe the variations in the performance of each architecture while also applying domain specific constrains that were designed to the visual and material features of Mesoamerican codices. Therefore, in the next chapter this research will analyze the training behavior of the main architecture CoRe-GAN, model to model comparisons across the previously mentioned experimental conditions, the mask specific performance analysis to observe how each architecture handles different degradation patterns and the statistical aggregation which includes mean, variance and visualizations of performance distributions.

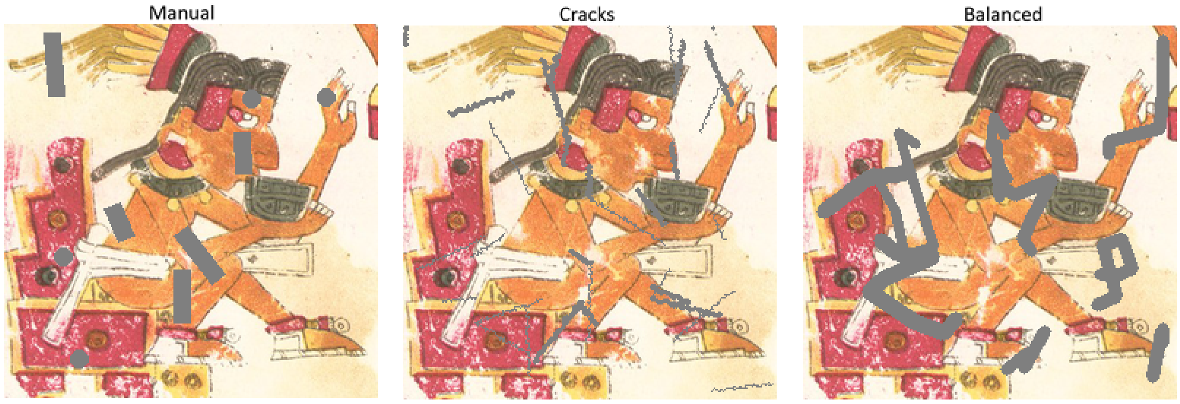

Figure 4: Mask types.

6 Evaluation

The following chapter provides an exhaustive evaluation of the CoRe-GAN system. First the internal learning dynamics of the model will be examined and then a quantitative and qualitative analysis of the results across all the model variants. The evaluation aims to reflect an structure that allows the clearly observation of the chronological and methodological progression of the study.

6.1 Training Dynamics and Model Progression

In order to fully understand and interpret the final performance of the model, it is required that there is an understanding of the internal behavior of the generator during the training. During this process, over 680 epochs were monitored, recording PSNR, SSIM, perceptual loss, LPIPS, generator loss and discriminator stability into a training log. These metrics allowed to supervise in real time the training process and also reveal how each training phase contributed to the development of the reconstruction capabilities of the model either on a structure, perceptual or texture level.

6.1.1 Pretraining: Learning Structure Without Adversarial Pressure

The pretrain phase ran for 60 epochs, and its main focus was on reconstruction driven objectives (L1, perceptual and LPIPS losses). This allows the generator to learn the rough semantic structure of the images without having to deal with adversarial pressure. The results from these stage shows that L1 loss decreased from $\sim$ 0.088 to $\sim$ 0.030, PSNR increased from $\sim$ 12 dB to $\sim$ 22 dB, SSIM also rose from $\sim$ 0.62 to $\sim$ 0.85, while LPIPS decreased from $\sim$ 0.21 to $\sim$ 0.11. These results indicate that even in the absence of a discriminator, the network was able to develop a strong foundation, catching the geometry and overall shapes that are present in the codex imagery. This shows a healthy early perceptual learning that will serve as a stable base for the GAN refinement.

6.1.2 Adversarial GAN Training: Enhancing Texture Realism

The PatchGAN discriminator introduced an adversarial feedback, which promotes the synthesis of realistic high frequency textures like granularity, contour transitions and characteristic strokes of the document. This stage consists of epochs 61-400. The metrics show that there was a significant reduction in adversarial loss oscillation after passing the ~ 50 epoch mark. Meanwhile, the LPIPS decreased steadily, which is an indicator of improvements in perceptual quality. PSNR plateaued around $23-26$dB, which is a common training behavior as the GAN optimization is prioritizing realism over pixel alignment. Finally, SSIM stabilized around 0.90-0.93. This phase is essential for the model because it captures the stylistic continuity of the Codex iconography. Pretraining learned what the missing structures should look like and GAN learned how the textures should be.

6.1.3 Five-Stage Fine-Tuning: Domain Adaptation for Codex Restoration

The final stage of the training was a fine tuning curriculum that was designed to enhance stability gradually, improve the sharpness at the edges, and that the network could learn degradation patterns that resemble real damage found in cultural manuscripts.

- Phase 1. Lower learning rate: it achieved a stabilized training, with reduced noise and a better metric consistency
- Phase 2. Reduced adversarial weight: it shifted the focus towards a better structural accuracy, which leads to the reduction of minor over smoothing elements.
- Phase 3. Increased L1 weight: this is were the most notable improvements occurred with L1 dropping from ~ 0.025 to ~ 0.019. This results into the edges and fine lines to be reconstructed more sharply.
- Phase 4. Softer masks: this enabled the model to better perform the reconstruction of the diffuse fading, erosion and abrasion marks. The best epoch of these 4 phases was saved as the finetune model variant.
- Phase 5. Codex-Specific damage: the phase named Codex100, and it trained the model using only the Codex like mask generator. This final stage resulted in the highest performance metrics with a PSNR of $29-30$dB, SSIM ~ 0.96, LPIPS ~ 0.037 and a L1 ~ 0.017 to ~ 0.018.

These values are higher than all the stages (Table 1) of the training and shows how the introduction of culturally informed synthetic deterioration was a cornerstone for achieving domain fidelity.

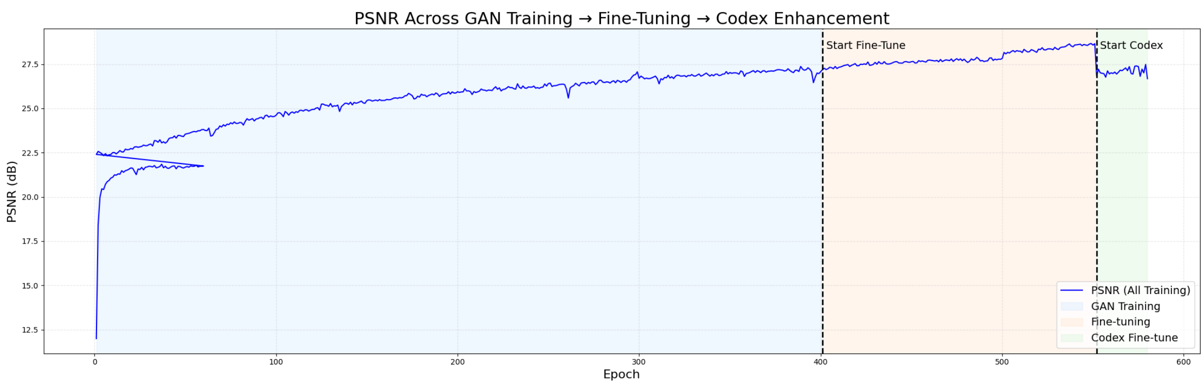

Figure 5: PSNR During CoRe-GAN training

Table 1: Summary of training behavior across model phases. Metrics reported as mean values over each stage.

Training Stage	PSNR (dB)	SSIM	LPIPS	L1 Loss
Pretraining (1–60)	12–22 (rising)	0.62–0.85	0.21–0.11	0.088 → 0.030
GAN Training (61–300)	Stable ∼ 23–26	0.88–0.93	0.11–0.06	0.030–0.025
Fine-tuning Stages 1–4	25–28 (gradual)	0.93–0.96	0.06–0.04	0.025–0.019
Codex100 Stage (Final)	29–30 (peak)	0.96	0.037	0.017–0.018

Table 2: Quantitative comparison of all evaluated models on the Codex Borgia restoration dataset. Values represent mean ± standard deviation across all test images. Higher is better for PSNR and SSIM; lower is better for LPIPS, MAE, and L1.

Model	PSNR	SSIM	LPIPS	MAE	L1
Codex100	29.483 ± 3.939	0.960 ± 0.026	0.037 ± 0.024	0.024 ± 0.010	0.024 ± 0.010
Finetune	29.469 ± 3.742	0.960 ± 0.023	0.038 ± 0.021	0.024 ± 0.009	0.024 ± 0.009
U-Net PatchGAN	23.711 ± 2.259	0.889 ± 0.037	0.117 ± 0.046	0.069 ± 0.014	0.069 ± 0.014
Context Encoder	15.525 ± 1.284	0.438 ± 0.077	0.532 ± 0.035	0.237 ± 0.039	0.237 ± 0.039

6.2 Model Comparison

While previous section analyzed how the CoRe-GAN model developed over time, this section evaluated the final performance of all the models, U-Net PatchGAN, Context Encoder, on the same unseen inference dataset, of 25 images. The inference applies standardized mask protocols and identical metrics. This results in to a fair comparison, controlled and independent of the training development of the baseline models. The results were logged into structured CSV files, which were later combined into an unified evaluation dataset that contains all the evaluation metrics for each model and mask combination. The unified evaluation gives us an insight into the models performances(Fig.6).

As we can observe in Table 2, the CoRe-GAN variant of Codex100 performs best across all the metrics, which points out how domain specific mask training highly enhances reconstruction quality. The CoRe-GAN variant Finetune is really close in performance, this is a sign that the gated U-Net architecture is stable, however finetune scored worse on LPIPS and PSNR, which indicates that the Codex100 benefited from culturally informed mask foundations. Meanwhile, UNet-PatchGAN has a reasonable structure recovery, however, it lacks texture fidelity. Finally, ContextGAN fails significantly in all the metrics, which points out the disadvantage against contextual attention and gating mechanisms.

6.2.1 Pairwise Improvements Analysis

This analisis shows how much the Codex100 model outperforms the baseline architectures across perceptual, structural and pixel level metrics. As we can observe in Table 3, the results shows that Codex100 provides a substantial gain when compared to earlier baselines, most specifically those not trained with codex like masks. Compared with the generic Context Encoder model, CoRe-GAN variant Codex100 achieves 89-120% improvement across all the metrics. This highlights how non specialized approaches when applied to historic documents have a clear disadvantage. Against the more modern model of U-Net PatchGAN, Codex100 yields 24% higher PSNR and a reduction between 65-68% of error in MAE, L1 and LPIPS.

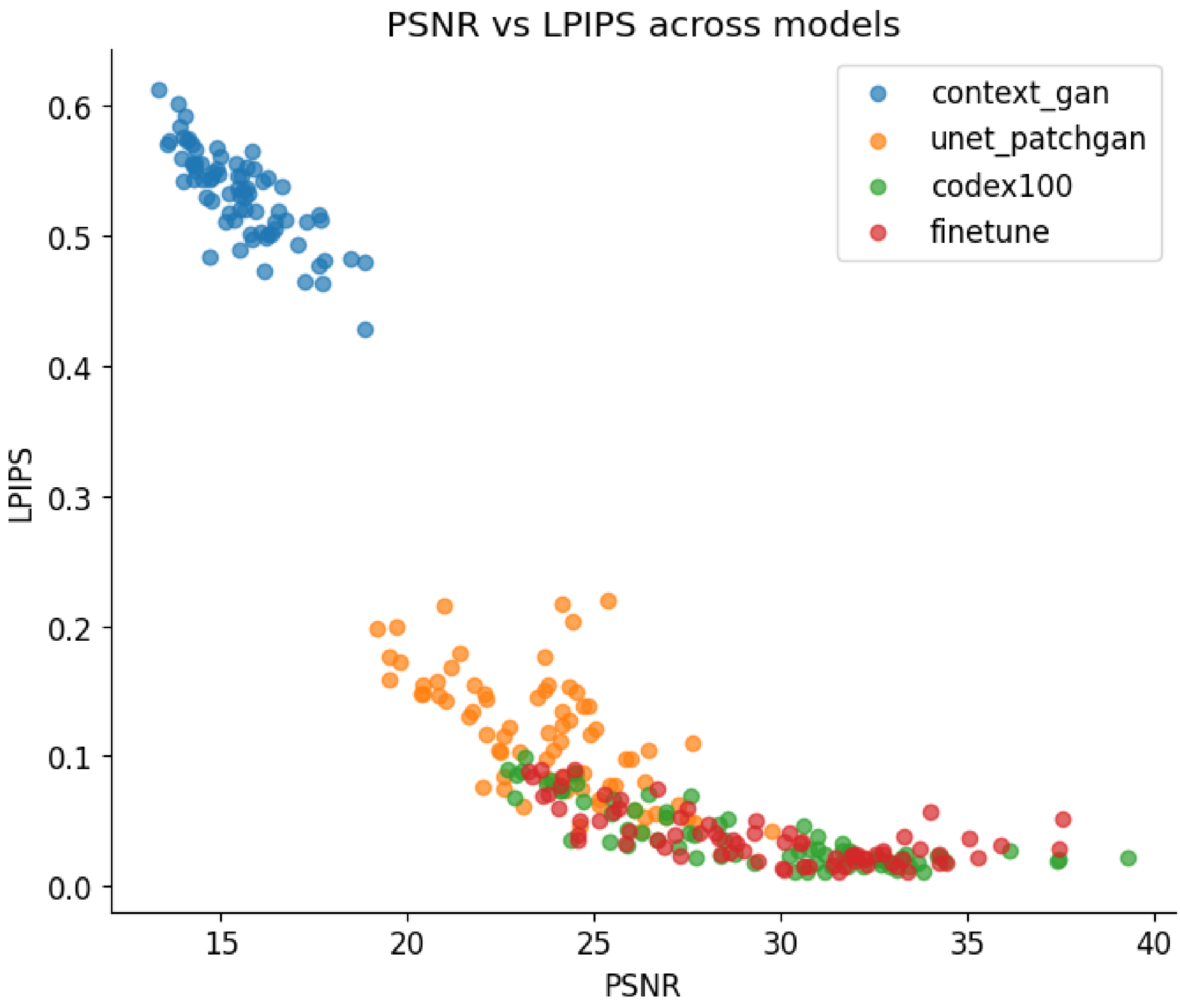

Figure 6: PSNR and LPIPS results per model.

On the other hand, when compared with the other variant of the CoRe-GAN, Finetune, the differences are marginal, less than 3%, which is expected as both share a gated U-Net architecture. The near scores show the stability and convergence of the CoRe-GAN architecture before the codex specific training. The massive improvements over the baselines demonstrate how domain aware synthetic masks and fine tune degradation modeling are the cause of those performance gains. In general, Codex100 variant has the best results in reconstruction quality, perceptual realism and structural fidelity, giving insight on how the cultural heritage specific training strategies in restoration pipelines play an essential role.

6.3 Discussion

This section reflects on the results presented in the previous chapter. The aim is to give them a significance for the restoration of Mesoamerican manuscripts, with an specific emphasis on the Codex Borgia. The progression contained in this investigation, from baseline architectures to refined variants give a clear example of methodological improvement. The findings also have remarks the implications for the design of deep learning systems meant to aid on the reconstruction of iconography of historical value.

Table 3: Percentage improvement of Codex100 over other models. Green indicates improvement, yellow indicates negligible change, and red indicates worse performance. For LPIPS, MAE, and L1 lower values are better.

Comparison	PSNR	SSIM	LPIPS	MAE	L1
Codex100 vs. Finetune	+0.047%	+0.00%	+2.63%	+0.00%	+0.00%
Codex100 vs. U-Net PatchGAN	+24.4%	+8.0%	+68.4%	+65.2%	+65.2%
Codex100 vs. Context Encoder	+89.8%	+119.4%	+93.0%	+89.9%	+89.9%

6.3.1 Interpretation of Training Progression

The training behavior that was documented in Section 6.1 shows that fine tuning by stages applied to the Gated U-Net with contextual attention architecture was essential to obtain the high quality restorations. More over, the pretraining phase allowed the generator to acquire a stable structural foundation, which resulted in the reduction of early reconstruction error in more than a half and improving the overall metrics. This can indicate that GAN-based inpainting architectures benefit from an initial phase where perceptual and reconstruction goals are prioritized before applying adversarial pressure.

The later GAN training gave improvements in the perceptual quality and the realism of the textures. These results highlights how an adversarial supervision for restoring complex patterns proper from Mesoamerican manuscripts is needed. If we skipped the the adversarial refinement,the model would have remained structurally accurate but would have lacked the high quality details that give makes a reconstruction look authentic.

Finally, fine tuning was the most significant stage in the model training. When introducing specific degradation masks with the CodexDamagaMaskGenerator, the network was able to improve in restoring real like deterioration. The overall best performance obtained by Codex1000 variant, suggests that specialized damage simulation is required to obtain a restoration that respects the cultural aesthetics of the codices. These findings nurture the field regarding on how synthetic damage should be modeled when training with a source of cultural heritage nature.

6.3.2 Comparative Model Performance and its Implications

When evaluating the performance across the models, it is evident that there are different capabilities between the baseline architectures and the CoRe-GAN (Fig.7). The Context GAN is missing attention to the context mechanisms and has a rather simple generator-discriminator structure, this is reflected when observing its results in pixel and perceptual level metrics. The overall performance of the Context GAN suggests that a generic context encoder is not enough when dealing with complex tasks of restoring intricated iconography.

Meanwhile, the UNet-PatchGAN architecture baseline performed in a moderate way. It was able to reconstruct the general structure but it was failing when reproducing texture that seemed real or in keeping the stylistic continuity. This is but a reflection of the limitations of UNet architectures when they are not paired with a perceptual criteria or a mask aware modulation of its features. In other words, it is able to preserve the geometry but it is not suited to produce a high fidelity synthesis of a material that is domain specific.

The Finetune model, which was trained with generic irregular masks, had a strong performance and represented an essential intermediate step in the way of the development pipeline. When compared to the later Codex100, some of its limitations where more visible. For example, Codex100 was trained with domain specific masks, and this yielded a lower LPIPS and L1, which point to a domain specific knowledge gives an edge over other models. These findings show how important is to tailor the training objectives and data representations to meet the needs of the cultural artifact that is being studied.

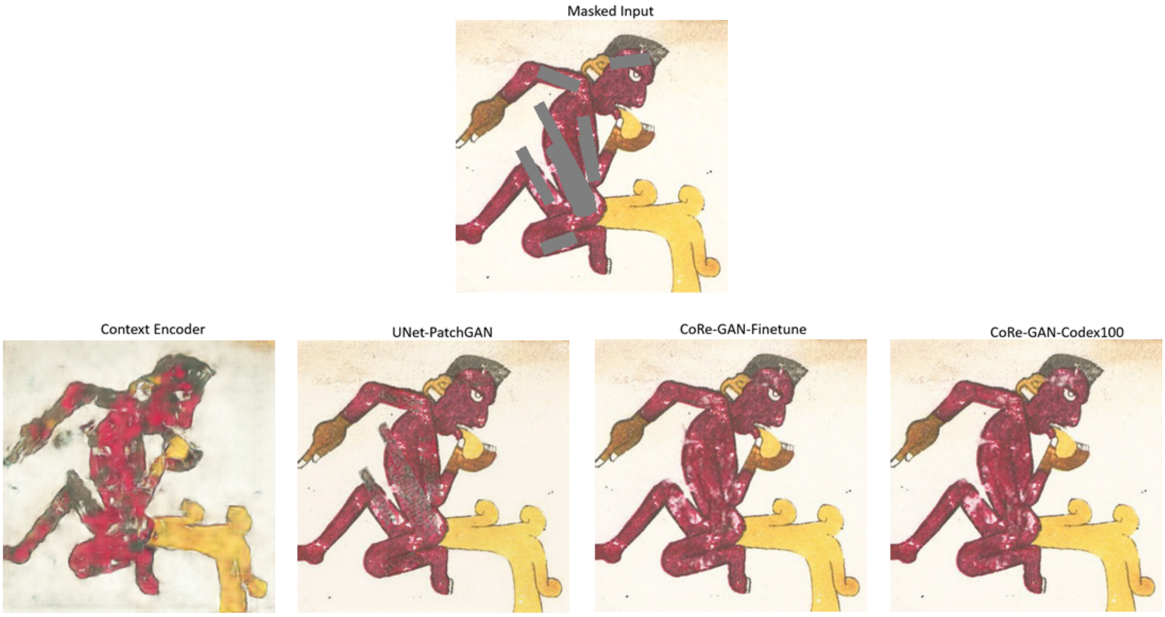

Figure 7: Output comparison.

6.3.3 Importance of Culturally Informed Mask Design

One of the most notable results of this research is the demonstration that a culturally informed synthetic damage simulation leads to an improved restoration. As stated before, the Codex images suffer from damage not often seen in other type of media. With that in mind, the CodexDamageMaskGenerator (Fig.8, 9) was designed, to emulate the effects of natural deterioration, and the model consequence of this approach, performed better than those trained with standard free form masks.

This result could have interdisciplinary implications for digital humanities and cultural heritage restoration research. It suggest that the training environment should mirror the physical processes that caused the damage that is aimed to restore. The domain specific foundations allows a model to recover some details that otherwise would not match visually with the historical techniques or the material characteristics. This approach likely extends when restoring ancient manuscripts, murals and frescoes, objects immersed in a cultural context where their degradation patterns mirrors specific environmental or material histories.

6.3.4 Methodological Contributions and Critical Reflections

The present study demonstrates the effectiveness of combining gated convolutions, contextual attention mechanisms, perceptual losses and adversarial training along with a

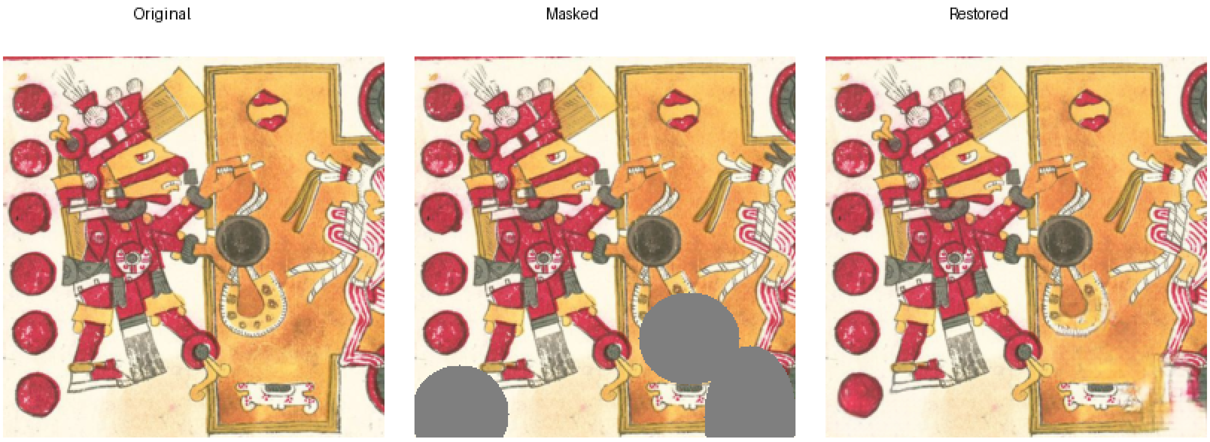

Figure 8: Codex Like Mask.

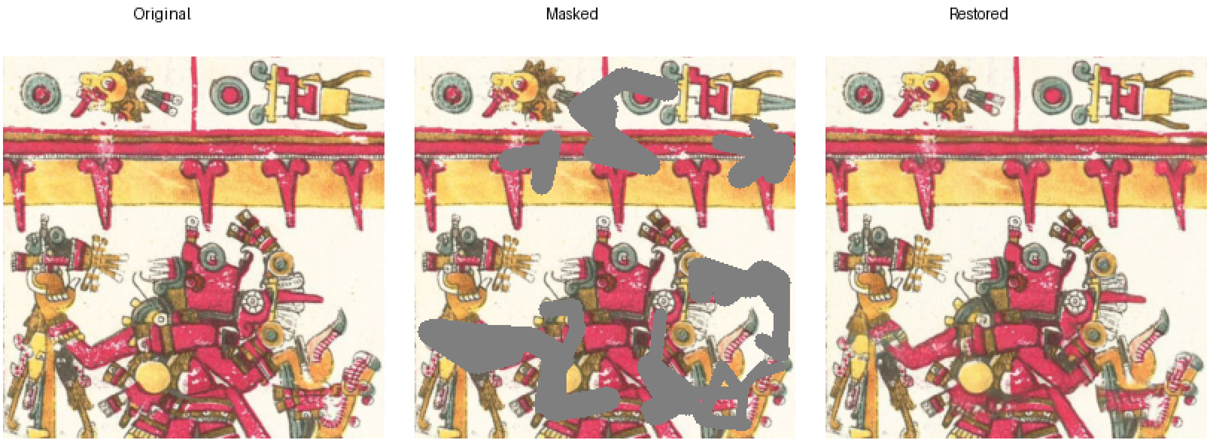

Figure 9: Free Form Mask.

multi phase curriculum, specially designed with the goal of restoring a cultural artifact. As the study analyzed the training journey and the final comparisons with the other models, it was observable that all of the components aided to the final performance of CoRe-GAN. The gated convolutions improved the mask awareness, while the contextual attention increased the spatial borrowing and the perceptual losses boosted the texture realism. The step by step design of the training process improved the stability and specialization of the model.

However, the methodological decisions that were taken also expose some limitations. The earlier baselines were not trained with the codex specific masks, which later found to be very beneficial. Although this does not compromise the comparisons between the models during the inference stage, it does limit the ability to generalize the conclusions about the usage of mask types in different architectures. In addition to this, the use of 256 x 256 images, chosen for computational feasibility due to limited access to the desired power, meant a inherent constraint on the model's ability to reproduce extremely fine line work or pigment granularity. The use of higher resolutions or hierarchical multi scale approaches could solve this limitation.

More over, even when the CodexDamagedMaskGenerator simulates the deterioration patterns present in the Codex Borgia, it is still a synthetic approximation. Some real world damage caused by various factors such as humidity exposure, biological mechanisms, conservation interventions or pigment chemistry where not addressed by this project. Future work could use the integration of physically based models or multi spectral imaging data to improve the authenticity.

6.3.5 Relationship to Prior Works

The findings of this research coincides with previous work made on GAN-based inpainting and cultural heritage restoration, but also extends the field of study. The research is similar with Ýu et al. (Yu et al.; 2018, 2019) about gated convolutions and contextual attention being crucial for high quality restorations. Studies over frescoes and mural restoration (Zhao et al.; 2024; Zhang et al.; 2025; Li et al.; 2024) emphasize the importance of using a domain specific mask design and applying a perceptual supervision. This project contributes to the literature by applying these strategies to Mesoamerican codices, a domain unique in its kind with an own visual structures and degradation patterns, and also by empirically demonstrating the advantages of using culturally informed synthetic masks in the training.

Through the evaluation section, this study positions CoRe-GAN, specially its Codex100 variant, as a specialized domain aware model custom designed to meet the challenges of ancient manuscript reconstruction.

7 Conclusion and Future Work

This study examined how a gated convolutional generative model can be adapted to restore damaged parts of the Codex Borgia, without compromising its original structure and style. In order to tackle gaps in existing inpainting models and cultural heritage restoration research, the project established four main objectives: creation of a curated Codex Borgia dataset, design of a domain-adapted restoration architecture, development of culturally informed deterioration masks, and rigorous evaluation against baseline models. All objectives were successfully achieved.

The proposed CoRe-GAN framework achieved strong restoration capabilities. The introduction of the Codex like mask generator, CodexDamageMaskGenerator became a central piece of to model performance, allowing to learn closer to real life deterioration conditions. Quantitative and qualitative evaluations showed that the final CoRe-GAN variant (Codex100) was the best performing model across all metrics in PSNR, SSIM, LPIPS, MAE, and L1 metrics, which shows the importance of domain specific mask design and perceptual supervision. The analysis of training dynamics gave an insight into how each phase contributed to structural accuracy and high-fidelity texture synthesis.

The study also acknowledges limitations. The 256×256 resolution restricts recovery of fine details, and synthetic masks cannot fully reproduce the complexity of physical degradation. Baseline comparisons were fair but constrained by training mask differences. However, the project delivers a novel dataset, a culturally informed restoration pipeline, and empirical evidence supporting the use of domain-specific generative models for manuscript restoration.

7.0.1 Future Work

There are directions that can develop this research further: applying a higher resolution or multiscale models to better capture fine line work and granularity. More realistic damage simulation by different techniques like multi spectral references. Feed the dataset with different codices to improve generalization across Mesoamerican documents. Interactive restoration tools that allows conservators to guide model outputs. All of these extension would improve and expand the models reliability and practical applications in cultural heritage conservation.

References

Ansari, M. (2025). Improving fine arts through artificial intelligence in art analysis, production, and restoration, *2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA)*, IEEE, pp. 1–7.
URL: *https://ieeexplore.ieee.org/document/11013667/*

Bertalmio, M., Sapiro, G., Caselles, V. and Ballester, C. (2000). Image inpainting, *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., USA, p. 417–424.
URL: *https://doi.org/10.1145/344779.344972*

Buvaneshwaran, B., Ashwin, V., Abishekrupan, B. S., Sasidharan, S. and Revathi, K. P. (2025). Ai driven restoration of damaged paintings and historical artifacts using generative adversarial networks, *ICDT 2025 - 3rd International Conference on Disruptive Technologies*, Institute of Electrical and Electronics Engineers Inc., pp. 1021–1025.

Criminisi, A., Perez, P. and Toyama, K. (2004). Region filling and object removal by exemplar-based image inpainting, *IEEE Transactions on Image Processing* **13**(9): 1200–1212.

Huang, S. and Hong, L. (2023). Diffusion model for mural image inpainting, *ITOEC 2023 - IEEE 7th Information Technology and Mechatronics Engineering Conference*, Institute of Electrical and Electronics Engineers Inc., pp. 886–890.

Iizuka, S., Simo-Serra, E. and Ishikawa, H. (2017). Globally and locally consistent image completion, *ACM Trans. Graph.* **36**(4).
URL: *https://doi.org/10.1145/3072959.3073659*

Isola, P., Zhu, J. Y., Zhou, T. and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks, *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Vol. 2017-January, Institute of Electrical and Electronics Engineers Inc., pp. 5967–5976.

Johnson, J., Alahi, A. and Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution, *Technical report*.

Kucuk, O. F., Bar, N. F. and Karakose, M. (2024). The role of image inpainting techniques in the restoration of artworks: An evaluation and comparative study, *2024 2nd International Conference on Sustaining Heritage: Embracing Technological Advancements, ICSH 2024*, Institute of Electrical and Electronics Engineers Inc., pp. 51–54.

Leong, W. Y. (2025a). Digital reconstruction of historical artworks using artificial intelligence, *2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI)*, IEEE, pp. 623–628.
URL: *https://ieeexplore.ieee.org/document/11020126/*

Leong, W. Y. (2025b). Revitalizing historical paintings through artificial intelligence, *10th International Conference on Digital Arts, Media and Technology, DAMT 2025 and 8th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, NCON 2025*, Institute of Electrical and Electronics Engineers Inc., pp. 141–145.

Li, L., Zou, Q., Zhang, F., Yu, H., Chen, L., Song, C., Huang, X., Wang, X. and Li, Q. (2024). Line drawing guided progressive inpainting of mural damage.
URL: *http://arxiv.org/abs/2211.06649*

Li, R., Pan, J., Li, Z. and Tang, J. (2018). Single image dehazing via conditional generative adversarial network, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, pp. 8202–8211.

Liu, X., Wan, J., Wang, N. and Wang, Y. (2024). Ancient painting inpainting based on multi-layer feature enhancement and frequency perception, *Electronics (Switzerland)* **13**.

Merizzi, F., Saillard, P., Acquier, O., Morotti, E., Piccolomini, E. L., Calatroni, L. and Dessì, R. M. (2024). Deep image prior inpainting of ancient frescoes in the mediterranean alpine arc, *Heritage Science* **12**.

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T. and Efros, A. A. (2016). Context encoders: Feature learning by inpainting, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544.

Telea, A. (2004). An image inpainting technique based on the fast marching method, *Journal of Graphics Tools* **9**.

Xu, Z., Zhang, X., Chen, W., Liu, J., Xu, T. and Wang, Z. (2024). Muraldiff: Diffusion for ancient murals restoration on large-scale pre-training, *IEEE Transactions on Emerging Topics in Computational Intelligence* **8**: 2169–2181.

Yan, Y., Chai, B. and Li, J. (2025). Artdiff: Integrating iot and ai to enhance precision in ancient mural restoration, *Alexandria Engineering Journal* **111**: 511–520.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. and Huang, T. (2019). Free-form image inpainting with gated convolution, *Technical report*.
URL: *https://github.com/JiahuiYu/*

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. and Huang, T. S. (2018). Generative image inpainting with contextual attention, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, pp. 5505–5514.

Zhang, J., Bai, S., Zeng, X., Liu, K. and Yuan, H. (2025). Supporting historic mural image inpainting by using coordinate attention aggregated transformations with u-net-based discriminator, *npj Heritage Science* **13**.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E. and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, pp. 586–595.

Zhao, F., Ren, H., Sun, K. and Zhu, X. (2024). Gan-based heterogeneous network for ancient mural restoration, *Heritage Science* **12**.