ADVANCED DEEP LEARNING TO GENERATE AND DETECT FAKE IMAGES OF EGYPTIAN MONUMENTS

by

DANIYAH ALASWAD

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL AND COMPUTER ENGINEERING

2025

Oakland University
Rochester, Michigan

Doctoral Advisory Committee:

Mohamed Zohdy, Ph.D., Chair
Subramaniam Ganesan, Ph.D.
Steven Louis, Ph.D.
Bill Solomonson, Ph.D.

To my mother and father

ACKNOWLEDGMENTS

This dissertation would not have been possible without the extraordinary support and love of those who stood by me throughout this journey. These words feel inadequate to express the depth of my gratitude.

To my parents: you planted the seeds of curiosity in me when I was just a child, and you have watered them with unconditional love and unwavering belief ever since. Every sacrifice you made, every word of encouragement, every moment you chose to invest in my dreams it all led to this. You have been the foundation upon which everything I am is built, and I carry your strength within me always.

Dr. Zohdy, my supervisor: Your mentorship reached into places where I needed it most offering comfort when anxiety threatened to overwhelm me, providing motivation when I couldn't find it within myself, and reminding me that the finish line, though distant, was always within reach.

Finally, to my siblings and friends: you have been my lifeline. You knew instinctively when I needed to laugh, to escape, to remember who I was beyond this work. You celebrated every small victory as if it were your own and held me together when setbacks threatened to break me. Your presence reminded me that life is richer and fuller than any single achievement, even when this dissertation consumed everything. You gave me perspective when I lost it and hope when it felt impossible to continue.
This achievement belongs to all of you as much as it does to me. You carried me here, and I will carry your love with me always.

Daniyah Alaswad

ABSTRACT

ADVANCED DEEP LEARNING TO GENERATE AND DETECT FAKE IMAGES OF EGYPTIAN MONUMENTS

by

Daniyah Alaswad

Adviser: Mohamed Zohdy, Ph. D.

This study examined the use of StyleGAN to create synthetic images of Egyptian monuments, addressing a critical gap at the intersection of generative artificial intelligence and cultural heritage. Through extensive experiments on a large dataset containing 5,000 Egyptian monument images, we show that architectural changes to the StyleGAN framework can significantly improve the quality and authenticity of the generated images. Our study contributes to the existing literature. First, we designed an enhanced discriminator architecture incorporating noise injection, squeeze-and-excitation blocks, and an improved MinibatchStdLayer, resulting in a Fréchet Inception Distance 27.5% better than that of the original model. We further introduced a novel image-text alignment approach using SigLIP, which can generate semantically guided monuments. We applied Differential Evolution (DE) to optimize the latent space of the conditional generator to reduce the alignment error by 15% for the targeted monument-generation tasks. We systematically analyzed various truncation methods used to manage noise in generated images by finding the best parameters that fit the architecture best but are also diverse. Statistical validation using bootstrap confidence intervals, McNemar's test and

DeLong's ROC analysis show significant improvements with effect sizes in the moderate to large range (Cohen's d ≈ 0.9-1.4) The discriminator was able to achieve 95.5% accuracy with a 5.3% false positive rate and 3.6% false negative rate. This 62% error drop was compared to the baseline. Under heavily corrupted conditions (JPEG quality = 10; Gaussian blur σ = 5.0), it achieved 78-85% of the baseline performance, whereas the default achieved 65-72% of the baseline performance. Frequency domain analysis results revealed resilience, with AUC values generally >0.95, varying by frequency. The new discriminator was approximately 20 to 25 percent more robust to adversarial attacks. However, both architectures are fundamentally vulnerable to stronger attacks. Our research shows how strategic refinements of operations models can produce representations of Egyptian monuments that attain a high-quality and satisfactory level of diversity that we can detect. Innovations can greatly help in the preservation of cultural heritage, virtual tourism, visualization, and education. This study will allow the generation of high-quality and varied Egyptian monument images, which can help in the digital conservation and easy accessibility of one of the world's great architectural heritages.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES—Continued

LIST OF ABBREVIATIONS

AUC	Area Under the Curve
AdaIN	Adaptive Instance Normalization
GAN	Generative Adversarial Network
FID	Fréchet Inception Distance
KID	Kernel Inception Distance
ReLU	Rectified Linear Unit
WGAN	Wasserstein Generative Adversarial Network
ProGAN	Progressive Generative Adversarial Network
SAGAN	Self-Attention Generative Adversarial Network
CLIP	Contrastive Language-Image Pretraining
StyleCLIP	StyleGAN + CLIP
SigLIP	Sigmoid Loss for Language-Image Pre-training
StyleGAN	Style Generative Adversarial Network
3D	Three-Dimensional
pp	Percentage Points
CNN	Convolutional Neural Network
HDF5	Hierarchical Data Format version 5
RGB	Red, Green, Blue (color space)
LR	Learning Rate
L2	L2 Regularization (Euclidean norm)
SNR	Signal-to-Noise Ratio

ROC	Receiver Operating Characteristic
LPIPS	Learned Perceptual Image Patch Similarity
IoU	Intersection over Union
PCA	Principal Component Analysis
BCE	Binary Cross-Entropy
IS	Inception Score

CHAPTER ONE

INTRODUCTION

This chapter explores the development and evaluation of a generative adversarial network model for creating realistic images of Egyptian monuments, detailing the techniques employed in this study.

3.2 Introduction

Generative Adversarial Networks (GANs) have revolutionized image synthesis, allowing the creation of highly realistic images in different domains [1]. Among these, StyleGAN has demonstrated state-of-the-art quality in the generation of high-resolution images through fine-grained style control [2]. The application of StyleGAN to generate high-fidelity synthetic images of Egyptian monuments has not been studied much, even though they are culturally and historically significant.

Goodfellow et al. [1] proposed that GANs consist of two neural networks: a generator and a discriminator. The generator produces images, and the discriminator differentiates between real and generated images. These neural networks continually enhance themselves and create increasingly realistic images through adversarial training [9]. GANs have been used for various tasks, such as image synthesis, super-resolution, and style transfer [10]. StyleGAN [2] extends this architecture with a style-based generator that maps random vectors to intermediate latent spaces and controls the image styles at different detail levels [2]. This allows for enhanced control of the visual characteristics to obtain the quality and diversity of the results.

Egyptian monumental architecture presents unique challenges for generative modeling due to its distinctive architectural elements, hieroglyphic ornamentation, weathered stone textures, and complex geometric proportions that span millennia of construction techniques. These buildings demonstrate that ancient Egypt had beautiful architecture and a strong culture. Monuments have severely deteriorated due to time, weathering, and various human activities. Advanced techniques can generate realistic images that are useful for preservation and virtual restoration [12].

Prior research has demonstrated that GANs can be used in a variety of heritage applications. For instance, for cultural heritage data augmentation, researchers have used GANs to enhance the training of computer-vision models [7]. Chinese calligraphy images were generated using StyleGAN [68]. The details and antique design style were captured satisfactorily. Nonetheless, limited research has focused on using StyleGAN to generate Egyptian monuments, leaving a gap in this important area of study [10].

Egypt is well-known for its ancient monuments, which attract millions of visitors annually [5]. However, their preservation faces challenges, including environmental factors, tourism pressure, and limited resources [6]. The advanced computer vision method StyleGAN offers significant potential for applications in virtual tourism, educational, and cultural heritage resources [15], [15].

This study investigated how StyleGAN can effectively generate realistic Egyptian monument images by examining the impact of different discriminator models and image manipulation techniques on the quality and realism. Chapter 2 presents a comprehensive literature review covering GANs, the StyleGAN architecture, and their cultural heritage applications. Chapter 3 details the methodology, including data collection, preprocessing,

model training, discriminator comparison. Chapter 4 describes image manipulation techniques and evaluation metrics. It also outlines the experimental setup and describes the SigLIP (Sigmoid Loss for Language-Image Pre-training) model and the Differential Evolution (DE) optimization approach implemented in this study. Chapter 5 presents a comprehensive performance evaluation using established metrics, including FID (Fréchet Inception Distance), KID (Kernel Inception Distance), and Precision-Recall, alongside rigorous statistical validation through bootstrap confidence intervals, McNemar's test, and DeLong's ROC analysis. Chapter 6 synthesizes the research findings, discusses their implications for both technical advancement and cultural heritage preservation, acknowledges the limitations, and proposes future research directions.

1.2 Problem Statement

Ancient Egyptian monuments are among the greatest architectural feats of humans. Nevertheless, they remain at risk due to environmental degradation, tourist pressures, and a lack of resources for development [6]. Despite the promising solutions offered by digital preservation, existing methods do not adequately portray the visual complexity and diversity of Egyptian monumental architecture [15].

Despite the demonstrated potential of GANs in cultural heritage applications, these applications have not yet addressed the challenges arising from Egyptian monuments, which are large in scale, precise in geometry, rich in ornamentation, and complex in weathering developed over millennia [3], [11]. This study aims to fill this gap by creating and assessing specialized StyleGAN architectures tailored for generating Egyptian monuments. This study investigates discriminator model variants that incorporate noise injection, squeeze-and-excitation (SE) attention blocks, and an

improved MinibatchStdLayer [4], [8]. It also considers image-text alignment techniques based on SigLIP [53]. In addition, it covers the discriminator statistical evaluation and noise control methods. The goal is to find effective ways to generate and detect historically accurate and visually compelling representations of Egyptian monuments in the future. The improved design of the discriminator showed significant improvements for each measure.

The importance of this study goes beyond technological innovation and is useful for applications in virtual tourism [122], digital conservation [16], and educational resource development [152]. Well-prepared high-quality images can help improve the global accessibility of Egypt's cultural heritage and serve as hardware for restoration efforts. The ability to produce realistic images of monuments with almost perfect discriminator performance can ensure trustworthiness for applications in heritage protection. This study will advance the technological toolkit for digital preservation and cultural heritage conservation efforts by proposing generative techniques for the visualization of monuments [12], [17].

1.3 Research Questions

The use of StyleGAN to generate realistic Egyptian monuments was investigated in this study. We focused on the optimization of the discriminator architecture. Moreover, other operations were performed on the images to improve the project outcome. The aim of this investigation was to ascertain whether generative adversarial networks can be adapted to different cultural heritage applications while ensuring architectural fidelity and generation diversity. The generation pipeline model covers the whole process from architectural design to statistical validation. In this study, research questions were based

on generation pipelines. Overall, five research questions – with included sub-questions – guided this study:

1. How does discriminator architecture influence StyleGAN's capability to generate authentic Egyptian monument images?
    - How can StyleGAN discriminator architecture be adapted to understand the architecture, hieroglyphics, and wear and tear common to Egyptian monuments?
    - What is the impact of architectural enhancements—including SE attention blocks, noise injection regularization, and improved minibatch statistical layers—on the quality and authenticity of generated monument images?
    - How well do hierarchical discriminator designs capture fine details (e.g., hieroglyphics, relief carvings) and global structures (e.g., columns and pyramids, temples) in Egyptian monuments?
    - Does the improved discriminator differentiate real photos from generated photos with greater accuracy than a baseline architecture?
2. How effectively can image-text alignment models optimize the semantic correspondence between textual descriptions and generated Egyptian monument imagery?
    - How does SigLIP compare with CLIP? Specifically, how well does SigLIP encode images, textual descriptions of Egyptian monuments, and different architectural styles and historical eras?
    - Which evaluation metrics best measure the semantic alignment between monument descriptions and generated images? We will be focusing on image generation using architectural terminologies and cultural heritage contexts.

- What strategies and techniques can be used for optimizing the textual description of the monument and the created image?

3. What optimization strategies most effectively navigate StyleGAN's latent space to generate Egyptian monuments aligned with specific architectural specifications?
    - How does the deployment of DE optimization techniques impact latent space exploration enabling the discovery of seed populations that generate monuments aligned with descriptions?
    - Which optimization method is best at exploring the high-dimensional latent space to find optimal seeds for a specific monument type, architectural period, or style?
4. How can noise control and truncation strategies optimize the trade-off between generation fidelity and architectural diversity for Egyptian monuments?
    - How do we find the appropriate truncation parameter range for Egyptian monument generation across monument types and stylistic periods to balance photorealistic fidelity and architectural diversity?
    - How do the different truncations of generated monuments affect their quality, architecture, and style?
    - What is the relationship between the noise control strategies, the architectural detail preservation, and the overall image realism of Egyptian monumental architecture?
    - How do systematic noise manipulation techniques affect the perceived authenticity and visual coherence of generated monument imagery?

5. What statistical evidence validates the performance improvements of the enhanced discriminator architecture compared to baseline StyleGAN3 (discussed later) implementations?
    - What improvements in various generation quality metrics, including FID, KID, IS (Inception Score), and Precision and Recall, does the enhanced discriminator achieve with respect to baseline architectures?
    - In what ways do bootstrap confidence intervals, McNemar's test, and DeLong's ROC analysis demonstrate statistical significance of observed performance differences between discriminator architectures?
    - What robustness improvements does the enhanced discriminator display across corruption types, adversarial perturbations, and frequency domain analysis?
    - How do the classification accuracy, error rates, AUC, and other metrics of the enhanced discriminator compared to baseline implementations in distinguishing authentic from generated monument images?

Research questions give us a systematic way towards adapting and optimizing StyleGAN for generating Egyptian monuments. Challenging issues are tackled at the Discriminator architecture design, Semantic alignment optimization, Latent space exploration, Quality-diversity trade-offs, and Statistical validation of the framework. The various bits of research carried out and presented in the paper are useful towards personalized generative modeling. Applications may span cultural heritage preservation, virtual reconstruction, and educational content creation.

1.4 Research Objectives

This study focused on leveraging StyleGAN to create photorealistic images of Egyptian monuments. A closer look was taken at architectural optimization and image manipulation techniques. Our research was guided by the following objectives:

- To create and assess a StyleGAN Framework specialized in producing Egyptian Monuments capable of generating visually appealing and culturally representative models that possess the distinctive features of these cultural heritage assets;
- To compare the impact of the discriminator architecture of StyleGAN on the quality of the generated images;
- To determine the best complex settings of the discriminator in order to capture the details of the decorative elements and the overall design of Egyptian monuments;
- To measure the performance of our architectural improvements using FID, KID, IS, and precision-recall curves, along with rigorous statistical validation using bootstrap confidence intervals, McNemar's test, and DeLong's ROC to quantify significance;
- To build and analyze SigLIP to create semantics between text and images through the example of Egyptian monuments;
- To compare the results of SigLIP with these standard CLIP models in terms of which one allows better control over the generated monument attributes;
- To implement DE for latent-space optimization in monument generation;
- To systematically study the effect of truncation parameters on the quality-diversity trade-off of the generated Egyptian monuments;
- To assess specialized techniques for reducing noise that maintain architectural details without introducing intrusive alterations;

- To develop guidelines for noise control to achieve maximum visual quality and perceived authenticity through the generation of monuments;
- To have a methodology that involves a robust evaluation against various corruption technique (e.g., JPEG compression, Gaussian blur, added noise, salt and pepper noise, and motion blur);
- To carry out a complete robustness evaluation across different quality evaluation dimensions: robustness against corruption/faults, frequency domain (multi-scale feature learning), semantic perturbation (lighting, style, perspective invariance), and adversarial robustness.

With these objectives in mind, we thoroughly studied the adaptation and optimization of StyleGAN for the generation and detection of Egyptian monuments.

1.5 Organization of Dissertation

The remainder of this dissertation is organized as a structured progression that guides the reader through the research process, findings and implications. Chapter 2 discusses the related work by establishing the research context in reviewing the relevant literature in primary domains. Chapter 3 describes the research methodology and describes the technical approach used in this study. Chapter 4 presents the implementation and evaluation and presents the comprehensive results and analyses. Chapter 5 presents the comprehensive statistical results, including the quantitative performance metrics. Finally, Chapter 6 presents the conclusions and future directions of this study. The final chapter synthesizes the research findings and discusses their implications.

CHAPTER TWO

LITERATURE REVIEW

2.1 Generative Adversarial Networks

2.1.1 Foundational Concepts

Generative Adversarial Networks (GANs) were introduced by Goodfellow et al. [1] as a new method for performing general modeling. The basic idea of GANs is based on adversarial training, where two neural networks (generator and discriminator) are trained in a minimax game. While the generator attempts to create synthetic data like the target distribution, the discriminator seeks to differentiate real samples from generated ones. This adversarial process is formalized as:

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}x \sim pdata(x)[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \qquad (2.1)$$

where G represents the generator, D the discriminator, $p_{\{data\}}$ the real data distribution, and p_z the prior distribution from which the latent vectors are sampled [1].

Karras et al. [3] highlighted that this adversarial training method has serious issues with mode collapse, instability during training, and upscaling of high-resolution images. Owing to these challenges, many architectural innovations have been introduced to improve GANs.

2.1.2 Architectural Innovations

Several GAN architectures have introduced novel and unique features to the domain. Deep Convolutional GANs (DCGANs), introduced by Radford et al. [5], implement convolutional layers and batch normalization with leaky rectified linear unit (ReLU) activations to improve quality and stabilize training. Gulrajani et al. [56] also

improved this method by adding some regularization with a gradient penalty. Karras et al. [3] proposed the progressive growth of GANs (ProGANs). This proposal aims to solve the problem of generating high-resolution images by adding layers to both the generator and discriminator during training. This progressive method allows stable training at higher resolutions than ever before. Zhang et al. [4] introduced self-attention GANs (SAGANs) that employed attention mechanisms to capture long-range dependencies in images to improve coherence in the generated content, notably for complex scenes with many objects. In this study, we applied deep learning-based techniques to the dynamic calibration of periodic or scrolling signals. We propose an automatic technique to calibrate a clock signal in an integrated circuit design.

2.1.3 Evaluation Metrics

The evaluation of GANs has evolved alongside architectural improvements. The IS quantifies the quality and diversity of the generated images. Salimans et al. [56] introduced this concept but noted the limitation of IS in capturing perceptual similarity with the training distribution. Heusel et al. [56] proposed the Fréchet Inception Distance (FID) – a more robust measure of generation quality – to compare the statistics of real and generated images in the feature space. Kynkäänniemi et al. [29] introduced the Precision and Recall metrics of GANs, which measure the quality and coverage of the learned distribution separately. These evaluation metrics are useful for improving the architecture and quantitatively comparing different GANs.

2.1.4 Statistical Validation Methodologies

Efron and Tibshirani [18] proposed the bootstrap method for estimating non-parametric confidence intervals. They also developed a Bca approach that corrects for

systematic bias and skewness. Such distributional properties are common in GAN evaluations [23], [57]. McNemar [19] developed paired statistical tests for assessing whether the error patterns of two classifiers differ while DeLong et al. [20] extended this analysis to test whether the ROC curves of two classifiers differ. Cohen [21] devised effect size measures (Cohen's d) to measure practical significance beyond p-values with d>0.8 considered large [21], [27], [54].

Hendrycks and Dietterich's systematic protocols for corruption tests include JPEG compression [44], Gaussian blur [46], additive noise [45], and motion blur. All of these tests simulate real-world situations for heritage photography [33], [43]. According to Kaneko and Harada [8], injecting noise during training helps maintain discriminative performance even when it is severely degraded. As examined by Wang et al. [148], efficient discriminators can take advantage of multi-scale frequency information [149]. Härkönen et al. [49] discovered that interpretable latent directions for semantic attributes can be obtained via GANSpace. Goodfellow et al. [39] proposed FGSM for adversarial attacks that was later expanded by Madry et al. [40] using iterative PGD. Carlini and Wagner [41] stressed the need for a broad attack evaluation process, which is especially important in authentication [38], [42].

2.2 StyleGAN Architecture

2.2.1 StyleGAN Development

StyleGAN, developed by Karras et al. [2], includes elements from the style transfer literature that make significant modifications to the GAN architecture. The most notable innovation of StyleGAN is the mapping network that converts the input latent code into an intermediate latent space that controls the styles at different resolutions

using adaptive instance normalization (AdaIN). Authors also redesigned the original architecture of StyleGAN to address its drawbacks: the generator architecture was redesigned to eliminate the appearance of water droplets, the path length was regularized to enhance the latent-space properties, and the progressive growth model was redesigned. These changes led to better-quality images and a disentangled latent space. Karras et al. [3] later proposed StyleGAN3, which aims to solve aliasing problems that lead to the handling of image artifacts through latent codes. StyleGAN3 features a generator architecture free of aliasing, which achieves strong equivariance, enabling image transformations without quality degradation.

2.2.2 Latent Space Properties and Manipulation

Owing to their unique properties, the latent spaces of StyleGAN models have received considerable attention. Shen et al. [48] showed that this latent space can be traversed in semantically meaningful ways so that images can be meaningfully controlled by finding directions and modifying them in the latent space. Härkönnen et al. [49] proposed techniques for latent-space factorization, enabling the disentanglement of interpretable directions in the latent space without explicit supervision of the model. These methods facilitate the control of generation actions. Heyi et al. [150 propose a method for projecting real images into the latent space of StyleGANs for editing real images. This capability can significantly aid in image restoration and enhancement.

2.2.3 Conditional Generation and Fine-tuning

Patashnik et al. [52] created StyleCLIP to blend StyleGAN with CLIP for image manipulation using text input. By utilizing this strategy, we can manipulate the meaning of the generated images using text. Zhang et al. [144] explored techniques to fine-tune

pre-trained StyleGAN models on limited data and showed that domain adaptation can be achieved using little data. This is especially true for more specialized domains, such as those involving cultural heritage, where there may not be much training data. Wu et al. [51] proposed StyleSpace analysis techniques that allow for finer control of visual properties in generated images, permitting more accurate manipulation of features relevant to domains.

2.3 GANs in Cultural Heritage Applications

2.3.1 Monument and Artwork Generation

Several studies have explored the application of GANs in cultural heritage. The unique features of a specific art style can be captured by image generation GANs. For example, GANs have been applied to generate historical art styles. Gao et al. [126] used GANs to generate realistic images of Chinese architectural heritage for application in architecture. Liu et al. [68] used StyleGAN to make complicated historical Chinese calligraphy images. The application shows that the model can learn fine-grained visual characteristics from the cultural domain. Sabatelli et al. [127] used GANs to generate art in specific, historical styles. They demonstrated that this generating ability is indeed possible. This study highlights how GANs can learn the visual grammar of culturally-specific architecture and artistic forms.

2.3.2 Digital Restoration and Reconstruction

Zhang et al. [145] employed GANs to implement a procedure for virtual restoration of damaged artworks. The authors reconstructed the missing parts of the artworks based on the remaining parts. This application could be useful for the digital conservation and visualization of damaged cultural heritage objects. Nogales [151] used

generative models to reconstruct architecture in cases where only partial remains were available. Their GAN-based method visualized the complete structure from archaeological fragments. This application is relevant to Egyptian monuments, many of which are in partially preserved conditions.

2.3.3 Virtual Access and Education

In reference, Mehta et al. [122] explored the use of GAN-generated imagery for virtual tour experiences of cultural heritage sites to help improve accessibility and engagement of global audiences with historic monuments. According to Mulders et al. [152], GANs can generate cultural heritage images for various teaching scenarios. The integration of GAN-generated imagery in education offers a great opportunity to teach history and architecture immersivity.

2.4 Egyptian Monuments in Digital Contexts

2.4.1 Digital Documentation and Preservation

Most existing studies on Egyptian monuments in the digital realm are documented and 3D reconstructed. YastiklI et al. [152] surveyed the use of photogrammetry and laser scanning to gather detailed information on Egyptian monuments for the establishment of digital baselines. Basu et al. [153] describe techniques for acquiring the lighting and material properties of Egyptian monuments so that they can be digitally reconstructed accurately. The data from this type of analysis can be used to train generative models that replicate the spatial characteristics of the structures.

2.4.2 Computer Vision Applications

Barucci et al. [11] used computer vision to classify and analyze hieroglyphics appearing on Egyptian monuments, highlighting the feasibility of performing automatic

visual analysis in archaeology. Hassan et al. [10] studied the techniques of image-based localization with respect to Egyptian archaeological sites, which pose various challenges to image recognition. The unique visual characteristics of Egyptian monuments have been understood from these findings.

2.4.3 Virtual Reconstruction

According to Pietroni et al. [12], the virtual reconstruction of partially destroyed Egyptian temples is based on archaeological data and procedural modeling techniques. This shows that even if GANs are not practically used, generative methods for archaeological visualization could be useful. Pietroni [12] developed methodologies for the evidence-based virtual reconstruction of Egyptian architectural heritage. The creation of GAN models for the creation of Egyptian monuments considers these factors.

2.5 Research Gaps and Opportunities

The literature identifies some crucial gaps and opportunities that can be useful for this study:

- Although StyleGAN architectures have proven their worth in several domains, including architectural heritage and Egyptian monuments, this has not yet been thoroughly investigated.
- Existing methods for the digitization of Egyptian monuments have only focused on documentation and reconstruction, not on generative modeling. Therefore, there is a huge potential.
- The unique appearance of Egyptian monuments, including their special architecture, hieroglyphics, and the materials they are made of, offers general modeling a challenge.

- The use of GAN-generated images of Egyptian monuments can have significant applications in cultural heritage, education, and virtual tourism.
- The technical contributions of optimizing the discriminator architecture of StyleGAN, implementing semantic control methods, and improving quality diversity, etc., in generating Egyptian monuments and architecture, are discussed.
- There is a lack of research on the discriminator part of StyleGAN.

These identified gaps reveal a critical need for domain-specific architectural adaptations of StyleGAN for cultural heritage applications, particularly for Egyptian monuments. While existing work demonstrates the feasibility of generative approaches in heritage contexts, no prior research has systematically addressed the unique challenges of Egyptian architectural generation with comprehensive discriminator evaluation and statistical validation. This dissertation addresses these gaps through developing and evaluating methods based on StyleGAN for generating Egyptian monuments with corresponding discriminator architectures, image-text alignment methods, optimization, and noise control methods.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

The chapter on the intersection of Egyptian monumental architecture and AI examines the role of Generative Adversarial Networks (GANs) in the visual generation of Egyptian monuments. The methodology deals with the implementation, training, and evaluation of StyleGAN models for creating realistic artwork representing Egypt. The center of this technique is the adversarial training process that characterizes GANs, where two neural networks engage in competitive learning dynamics. Over the years, the generator network improved the creation of synthetic images of Egyptian monuments. Meanwhile, the discriminator network improved the ability to distinguish photos from generated images. The competitive relationship forces both networks to improve their characteristics.

The methodology explains how StyleGAN was adapted to Egyptian monuments while also maintaining variety in architectural components and the quality of images. Through a range of parameter tuning and training techniques, including truncation optimization, we produced photos that were both realistic and architecturally authentic to the specifics of Egyptian monuments.

The 1818aluation methods (FID, PR, and IS) and the evaluation of architectural authenticity were important parts of this study. The evaluation methods allow for the knowledge that a model has regarding a specific image subject. The methodology incorporates ethical concerns in response to issues raised regarding representing Egypt

culturally, repeating history, and the abuse of AI images. This study explores the use of generative algorithms in AI image generation for the architectural heritage and historic monuments of Egypt.

3.2 Generative Adversarial Networks (GANs)

In 2014, Goodfellow et al. proposed Generative Adversarial Networks (GANs) for unsupervised machine learning. GANs utilize an adversarial framework between a generator and discriminator (two distinct neural networks), which is vastly different from traditional feed-forward algorithms [1].

4.2.2 Fundamental Architecture

The GAN architecture consists of two main components that work against each other.

1. A generator network synthesizes fake data (in our case, images of Egyptian monuments) from random noise vectors. It takes random latent space vectors and generates images resembling the statistics and appearance of actual Egyptian monuments.
2. The discriminator network is essentially a binary classifier that determines whether the input image belongs to the training dataset of real Egyptian monuments or is generated by the generator. The probability scores indicate the accuracy of the images.

3.2.2 Adversarial Training Dynamics

The competitive framework of the two networks is the key to the GAN concept. This relationship can be conceptualized as a minimax game:

- The generator attempts to make the discriminator think that its outputs are real so that the discriminator cannot distinguish them.
- The goal of the discriminator is to maximize the accuracy of distinguishing between real and generated images.

This adversarial system increases through continuous improvement:

1. As the discriminator becomes more adept at detecting synthetic images, it provides more precise feedback to the generator.
2. The generator uses this feedback to refine its outputs, producing increasingly convincing images of Egyptian monuments.
3. As the generated images improve, the discriminator must further enhance its detection capabilities.

As a result of this repetitive adversarial training, both the generator and discriminator improved over time. Eventually, the generator produced Egyptian monument images that maintained the original training samples' architecture, culture, and visuals.

3.2.3 Model Architecture

A typical GAN architecture comprises two fundamental neural network components: a generator and discriminator. In the initial training phase, both networks randomly initialize the parameters. The generator transforms random noise vectors into synthetic data, whereas the discriminator evaluates both authentic examples from the training dataset and the artificial data produced by the generator.

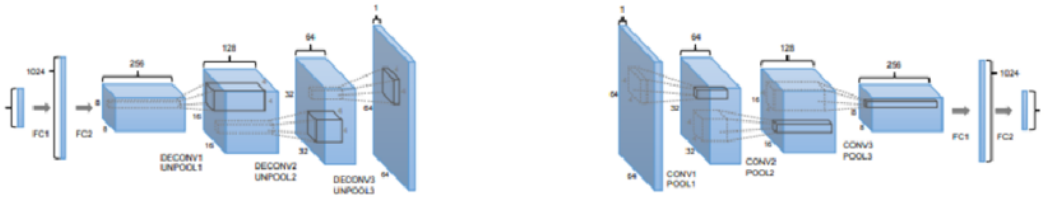

Figure 3.1: GAN

The GAN framework can be formalized as a minimax game where the generator G aims to minimize this objective while the discriminator D aims to maximize it, creating the adversarial dynamics central to GAN training, represented by the following value function [42] [43]:

$$\min_G \max_D V(D,G) = \mathbb{E}x \sim pdata(x)[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \quad (3.1)$$

where

G the generator function;

D the discriminator function;

$V(D,G)$ the value function representing the objective of the GAN;

x a real data sample from the training dataset;

z a random noise vector sampled from a prior distribution (e.g., Gaussian);

p_{data} the true data distribution;

$\mathbb{E}$ the expected value.

3.2.4 Training Phases

The training alternated between two different phases:

1. Discriminator Training: The discriminator learns to distinguish between real and fake data. It aims to minimize the classification errors. Specifically, it aims to:

- Identify real images as true (value 1)
- Identify generated images as fake (value of 0)

The discriminator optimizes its parameters to maximize $V(D, G)$, effectively maximizing the log probability of the correct classification of real and generated samples.

2. Generator Training: The generator is improved to generate data that could deceive the discriminators. Its objective is to:

- Create images that the discriminator thinks are real but are false
- Minimize the term $\mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right]$

In practice, to prevent gradient saturation during early training, the generator often maximizes $\mathbb{E}_{z\sim p_z(z)}\left[\log D\left(G(z)\right)\right]$ rather than minimizes the original term.

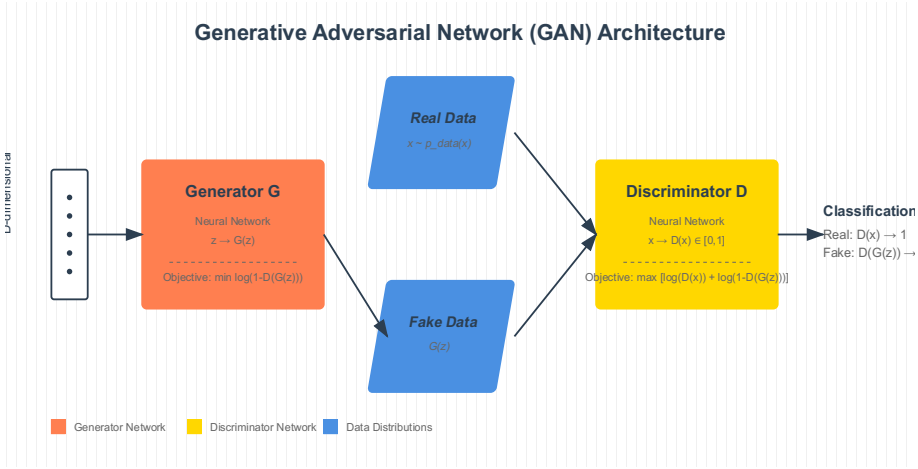

Figure 3.2: GAN Architecture

For generating Egyptian monuments, the discriminator D learns to differentiate between real architectural features like hierarchies, hieroglyphs, weathering, proportions, and artifacts from the generator.

3.2.5 Loss Function

The loss function in a machine learning model serves as an indicator of its performance. The predictive accuracy of the model was evaluated by comparing its predictions with actual outcomes. In the case of a generative adversarial network (GAN), the loss function typically comprises two components: one associated with the generator and the other with the discriminator.

The discriminator loss function measures the effectiveness of the discriminator in distinguishing between real Egyptian monument images and those generated by the generator model. It aims to maximize the probability assigned to real images while minimizing the probability assigned to the generated images. A well-performing discriminator produces values close to 1 for real monument images and values close to 0 for generated images, thus minimizing the loss function.

$$L_D = -Ex \sim pdata(x)[\log D\,(x)] - E_{z\sim p_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{3.2}$$

The generator loss function evaluates the success of the generator in creating images that deceive the discriminators. The generator aims to maximize the probability that the discriminator incorrectly classifies its output as authentic.

$$L_G = -E_{z\sim p_z(z)}\left[\log D\,\big(G(z)\big)\right] \tag{3.3}$$

This formulation encourages the generator to produce Egyptian monument images that the discriminator misclassifies as real images. The objective of the generator is to minimize the loss function. These two components create an adversarial dynamic:

- The discriminator tries to minimize L_D, improving its classification accuracy.
- The generator tries to minimize L_G, enhancing its ability to create convincing Egyptian monument images.

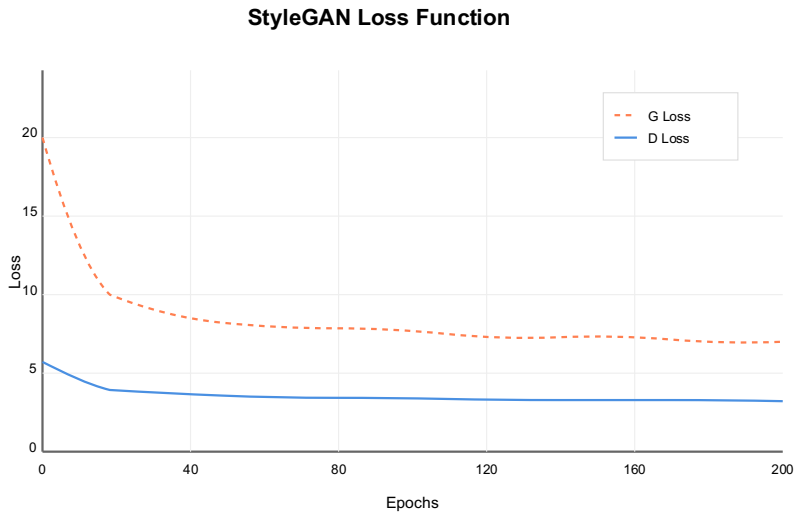

Figure 3.3: Loss Function Plot

The ultimate objective is to achieve equilibrium, where the generator creates Egyptian monument images such that the discriminator can perform no better than random guessing, effectively generating samples that are indistinguishable from authentic monument photographs [44], [45].

3.2.6 Generator

A generator in a GAN transforms random noise or input vectors into data sampled similar to the training data distribution of the Egyptian monument. The architectural structural process gradually builds complex visuals through transformation. The typical generator architecture consists of some important parts that act in the following order:

1. The process begins with a dense layer that maps random noise vectors (usually sampled from a Gaussian or uniform distribution) to a specific representation.
2. The second path of the network deals with a process known as upsampling. The transposed convolutional layers are arranged such that the network can

progressively upsample the low-dimensional latent representation into a high-dimensional image.

3. To stabilize training, normalization is usually followed by batch normalization in every transposed convolutional layer, which accelerates the process.
4. After normalizing the output, nonlinear activation functions, mainly rectified linear units and leakyReLUs, were employed.
5. The last layer uses the tanh or sigmoid activation function to provide a pixel value within the correct range for the image data.

Generator	Parameters	Buffers	Output shape	Datatype
mapping.fc0	262656	-	[16, 512]	float32
mapping.fc1	262656	-	[16, 512]	float32
mapping	-	512	[16, 16, 512]	float32
synthesis.input.affine	2052	-	[16, 4]	float32
synthesis.input	262144	1545	[16, 512, 36, 36]	float32
synthesis.L0_36_512.affine	262656	-	[16, 512]	float32
synthesis.L0_36_512	2359808	25	[16, 512, 36, 36]	float32
synthesis.L1_36_512.affine	262656	-	[16, 512]	float32
synthesis.L1_36_512	2359808	25	[16, 512, 36, 36]	float32
synthesis.L2_36_512.affine	262656	-	[16, 512]	float32
synthesis.L2_36_512	2359808	25	[16, 512, 36, 36]	float32
synthesis.L3_52_512.affine	262656	-	[16, 512]	float32
synthesis.L3_52_512	2359808	37	[16, 512, 52, 52]	float16
synthesis.L4_52_512.affine	262656	-	[16, 512]	float32
synthesis.L4_52_512	2359808	25	[16, 512, 52, 52]	float16
synthesis.L5_84_512.affine	262656	-	[16, 512]	float32
synthesis.L5_84_512	2359808	37	[16, 512, 84, 84]	float16
synthesis.L6_84_512.affine	262656	-	[16, 512]	float32
synthesis.L6_84_512	2359808	25	[16, 512, 84, 84]	float16
synthesis.L7_148_512.affine	262656	-	[16, 512]	float32
synthesis.L7_148_512	2359808	37	[16, 512, 148, 148]	float16
synthesis.L8_148_512.affine	262656	-	[16, 512]	float32
synthesis.L8_148_512	2359808	25	[16, 512, 148, 148]	float16
synthesis.L9_148_362.affine	262656	-	[16, 512]	float32
synthesis.L9_148_362	1668458	25	[16, 362, 148, 148]	float16
synthesis.L10_276_256.affine	185706	-	[16, 362]	float32
synthesis.L10_276_256	834304	37	[16, 256, 276, 276]	float16
synthesis.L11_276_181.affine	131328	-	[16, 256]	float32
synthesis.L11_276_181	417205	25	[16, 181, 276, 276]	float16
synthesis.L12_276_128.affine	92853	-	[16, 181]	float32
synthesis.L12_276_128	208640	25	[16, 128, 276, 276]	float16
synthesis.L13_256_128.affine	65664	-	[16, 128]	float32
synthesis.L13_256_128	147584	25	[16, 128, 256, 256]	float16
synthesis.L14_256_3.affine	65664	-	[16, 128]	float32
synthesis.L14_256_3	387	1	[16, 3, 256, 256]	float16
synthesis	-	-	[16, 3, 256, 256]	float32
---	---	---	---	---
Total	28472133	2456	-	-

Figure 3.4: Summary of the Generator

3.2.7 Discriminator

The discriminator is an important evaluative component of the GAN, which acts as an expert classifier to distinguish between real Egyptian monuments and those generated by the generator. This neural network was trained to classify the input data as real or artificial. The structure of the discriminator usually has a sequence of convolutions

that downsample spatially and upsample feature-wise. The standard implementation includes the following steps:

1. First is a series of convolutional operations which help identify whether an object exists in an image. The layers at the beginning capture simple features, such as edges and textures. More complex features, particularly those of Egyptian architecture, such as hieroglyphics, column shapes, and monumental scales, are found in deeper layers.
2. After the convolution operations, the batch-normalization layers stabilize the learning process by normalizing the activations. The reduction in the internal covariate shift allows the use of considerably higher learning nonlinear activation function (usually Leaky ReLU with a low negative slope). This adds nonlinearity to the model. This helps the model learn complex patterns and prevents the dying ReLU problem.
3. The four Dropout Layers randomly deactivate some neurons in the model during training to prevent overfitting and improve generalization.
4. In the flattening operation, the feature maps obtained from multiple convolutional layers are transformed into a one-dimensional vector.
5. Complex relationships between features are established using fully connected layers.
6. The output layer contains a single neuron and uses a sigmoid activation function. The output layer compresses the final decision into a probability between 0 and 1. This indicates whether the input image is real or fake.

During training, the discriminator alternates between analyzing real Egyptian monument images aiming to obtain output values close to 1 and evaluating generator-created images aiming to output values close to 0.

As training progresses, the discriminator becomes increasingly adept at spotting flaws in the generated images, including impossible spaces, lighting, and proportions specific to Egyptian monumental architecture. The generator relies on feedback from the discriminator for guidance and improvement, adapting to implicit suggestions regarding the unrealistic features and appearance of the generated Egyptian monuments. Through this feedback loop, the generator is gradually forced to generate increasingly convincing representations of Egyptian monumental architecture [47], [48], [49].

Discriminator	Parameters	Buffers	Output shape	Datatype
---	---	---	---	---
b256.fromrgb	512	16	[16, 128, 256, 256]	float16
b256.skip	32768	16	[16, 256, 128, 128]	float16
b256.conv0	147584	16	[16, 128, 256, 256]	float16
b256.conv1	295168	16	[16, 256, 128, 128]	float16
b256	-	16	[16, 256, 128, 128]	float16
b128.skip	131072	16	[16, 512, 64, 64]	float16
b128.conv0	590080	16	[16, 256, 128, 128]	float16
b128.conv1	1180160	16	[16, 512, 64, 64]	float16
b128	-	16	[16, 512, 64, 64]	float16
b64.skip	262144	16	[16, 512, 32, 32]	float16
b64.conv0	2359808	16	[16, 512, 64, 64]	float16
b64.conv1	2359808	16	[16, 512, 32, 32]	float16
b64	-	16	[16, 512, 32, 32]	float16
b32.skip	262144	16	[16, 512, 16, 16]	float16
b32.conv0	2359808	16	[16, 512, 32, 32]	float16
b32.conv1	2359808	16	[16, 512, 16, 16]	float16
b32	-	16	[16, 512, 16, 16]	float16
b16.skip	262144	16	[16, 512, 8, 8]	float32
b16.conv0	2359808	16	[16, 512, 16, 16]	float32
b16.conv1	2359808	16	[16, 512, 8, 8]	float32
b16	-	16	[16, 512, 8, 8]	float32
b8.skip	262144	16	[16, 512, 4, 4]	float32
b8.conv0	2359808	16	[16, 512, 8, 8]	float32
b8.conv1	2359808	16	[16, 512, 4, 4]	float32
b8	-	16	[16, 512, 4, 4]	float32
b4.mbstd	-	-	[16, 513, 4, 4]	float32
b4.conv	2364416	16	[16, 512, 4, 4]	float32
b4.fc	4194816	-	[16, 512]	float32
b4.out	513	-	[16, 1]	float32
---	---	---	---	---
Total	28864129	416	-	-

Figure 3.5: Summary of the Discriminator

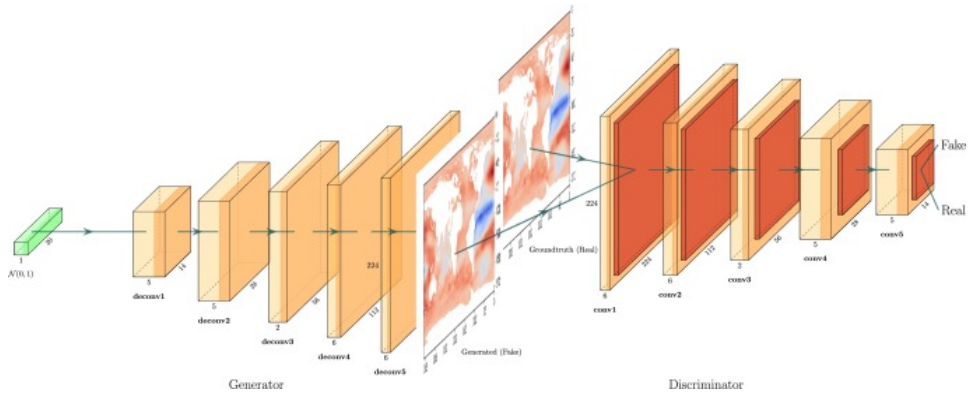

Figure 3.6: Model Architecture [148]

3.2.8 Evaluation and Iterative Refinement

Evaluating StyleGAN models for the generation of Egyptian monuments requires both quantitative and qualitative assessments of the generated outputs. The objective of this multi-pronged evaluation strategy is to iterate the process to improve the quality and diversity of the generated imagery of the monument.The main metric for comparing the distribution of generated images of Egyptian monuments to real ones will be the Fréchet Inception Distance (FID). The mathematical formulation is expressed as follows [3]:

$$\text{FID}(p, q) = \left|\left|\mu_p - \mu_q\right|\right|^2 + \text{Tr}\left(\Sigma_p + \Sigma_q - 2\left(\Sigma_p\Sigma_q\right)^{1/2}\right) \tag{3.4}$$

where

$\text{FID}(p, q)$	Fréchet Inception Distance between distributions p and q
μ_p and μ_q	mean feature vectors of distributions p and q respectively
Σ_p and Σ_q	covariance matrices of distributions p and q respectively
$\|\| \cdot \|\|^2$	squared Euclidean norm
$\text{Tr}(\cdot)$	the trace of a matrix
$\left(\Sigma_p\Sigma_q\right)^{1/2}$	the matrix square root

A lower FID score indicates that the generated monument distribution is more similar to the authentic distribution, and an FID score of 0 implies a perfect match. This metric is meaningful in terms of how well the model can capture the different architectural styles and features of Egyptian monuments [2].

The IS provides an additional evaluation of the monument image generation because it measures quality and diversity:

$$IS(G) = \exp\left(\mathbb{E}_{x\sim p_g}\left[D_{KL}\left(p(y \mid x) \parallel p(y)\right)\right]\right) \tag{3.5}$$

where

$IS(G)$ Inception Score of the generator G

x an image generated by the generator G

$p(y|x)$ conditional class distribution of the Inception classifier given image$x$

$p(y)$ marginal class distribution of the InceptionV3 classifier

$D_{\{KL\}}$ Kullback-Leibler divergence

$\mathbb{E}_{x\sim p_g}$ expectation over images generated by G

exp exponential function

Higher IS scores indicate that the image generated by the model is diverse and identifiable and can take multiple styles from different historic periods of the Egyptian monument.

The ability to precisely measure the coverage of the learned distribution, with an emphasis on quality, can provide a better evaluation of the generative model.

$$Precision = \frac{\#\ of\ generated\ samples\ that\ match\ real\ data\ distribution}{total\ \#\ of\ generated\ samples} \tag{3.6}$$

$$Recall = \frac{\#\ of\ real\ data\ distribution\ samples\ captured\ by\ getnerator}{total\ \#\ of\ real\ data\ distribution\ samples} \tag{3.7}$$

In practice, these values are approximated using nearest-neighbor calculations in feature spaces. This evaluation method is useful for identifying situations in which a model can have an excellent FID score but still suffers from low recall (high precision but mode collapse) or generates diverse samples that have low quality (high recall but low precision).

Visual inspection occurs during the early stages of training, when quantitative metrics do not represent the perceptual quality of the images. As evidenced by the

example images from the initial epochs of training, the model has yet to learn what a coherent-looking monument structure is like. By this exhaustive evaluation strategy, the iterative refinement process enhances the model's capability to generate true, diverse, and historically accurate representations of Egyptian monuments [3].

Data Collection and Preparation

This study used a combination of different complementary methods to create a composite database of the Egyptian monuments. The primary data sources included the following: The primary data source was the Egypt Monuments Dataset (222MB; available on Kaggle); this source provided a curated base for the monument images. The key features of the collected dataset are as follows:

- The dataset size was approximately 5,000.
- Different formats were standardized in the preprocessing step; resolution was standardized to 256×256 during the preprocessing of the inputs.
- There was a diverse inclusion of ethnicities, cultures, and countries in a global content context.

- Images includes pyramids, temples, obelisks, sphinxes, and other architectural elements of Egypt.

3.3.1 Ethical usage

Several measures were taken to ensure the quality and ethical use of the dataset:

- We developed a method to remove duplicate images in any format; this ensures that the standard is maintained. The images will be inspected.
- Public domain verification was performed.
- Data was handled ethically, according of the various data collection platforms.

The dataset is so comprehensive that it comes from multiple sources and pertains to different types of monuments in Egypt. Thus, it will provide an excellent training set for the StyleGAN model to generate authentic images of monuments.

3.3.2 Color Space Normalization

We removed the alpha channels where applicable and converted all images to RGB in order to standardize the colors of the dataset. For images with alpha channels, this involved either alpha compositing against a white background or direct channel removal from the image.

$$I_{\text{RGB}} = f_{\text{conversion}}(I_{\text{original}}) \tag{3.8}$$

3.3.3 Dimensional Standardization

The bilinear interpolation method was used to resize all images to 256 × 256 pixels. The resizing operation can be mathematically formulated as:

$$I_{\text{resized}}(x', y') = \sum_{i=0}^{1} \sum_{j=0}^{1} I_{\text{original}}(\lfloor x \rfloor + i, \lfloor y \rfloor + j) \cdot w(i, j, \alpha, \beta) \tag{3.9}$$

where

$I_{\text{original}}, I_{\text{resized}}$ original, resized images;

$w(i, j, \alpha, \beta)$ bilinear interpolation weights;

$w(i, j, \alpha, \beta)$ $= (1-\alpha)^{1-i} \cdot \alpha^{i} \cdot (1-\beta)^{1-j} \cdot \beta^{j}$;

α, β fractional parts of the source coordinates;

α $= x - \lfloor x \rfloor$;

β $= y - \lfloor y \rfloor$;

x, y source coordinates

$(x', y'): x$ $= x' \cdot S_x$,

y $= y' \cdot S_y$

S_x, S_y scaling factors in x and y directions

3.3.4 Pixel Intensity Normalization

To ensure that the gradient behaved consistently during training, the pixel values were normalized to between 0 and 1.

$$I_{\text{normalized}}(x, y) = \frac{I_{\text{resized}}(x,y)}{255} \tag{3.10}$$

3.3.5 Data Storage Format

The standardized images were converted to the TFRecords format, a binary storage format useful for handling training data in TensorFlow-based models such as StyleGAN3. This format has several advantages:

1. Efficient serialized binary storage is supported.
2. Optimized I/O performance for training purposes.
3. Preprocessing overhead during training iterations is reduced.
4. Parallel data loading and prefetching is supported.

The training process was bottlenecked owing to the slow read speeds of the datasets, which was solved by using TFRecords. A preprocessing pipeline was designed and developed to allow the StyleGAN model to receive consistent inputs. This is said to influence the training dynamics and convergence [3].

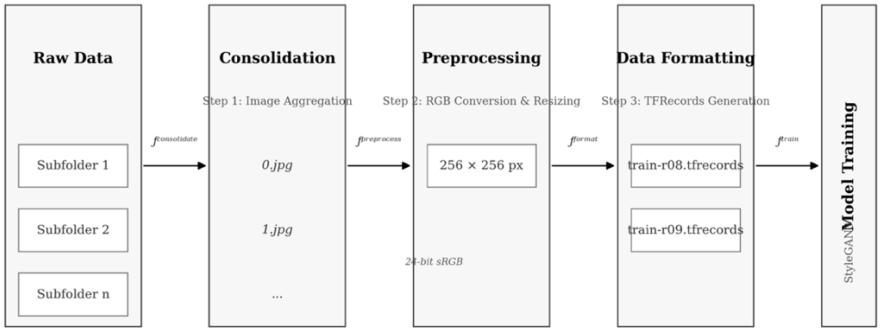

Figure 3.8: StyleGAN Data-processing Pipeline

3.4 Training Configuration

To determine whether the improved StyleGAN model could produce images of Egyptian monuments, experiments were conducted. The model was trained on a specialized dataset to improve the generation of Egyptian landmarks and enhance the image quality. The StyleGAN model, pretrained on an extensive dataset, used its prior learning to improve certain details of Egyptian monuments. We were able to manipulate facial expressions, pyramids, and architectural details. [2].

The experimental process involved fine-tuning of hyperparameters, implementation of progressive growth techniques, data augmentation, and calculation of performance metrics to evaluate the quality, diversity, and authenticity of the generated Egyptian monument images. This study aimed to indicate, through qualitative and

quantitative evidence, that the StyleGAN model generates convincing images of Egyptian monuments that closely resemble photographs of actual Egyptian monuments.

3.4.1 Training Configuration

The following configurations were used for fine-tuning:

- Base Model: stylegan3-t.
- The model was trained for a total of 25,000 iterations (approximately 7 days on 2 NVIDIA A100 GPUs (40GB memory each)).
- The batch size was set to 16 and the learning rate to 0.002 with adaptive discriminator augmentation strength.

3.4.2 StyleGAN Generator and Discriminator Training Process

The process of training StyleGANs is like that of GANs; the framework is adversarial and consists of a generator and discriminator. The training methodology is formalized in this overview.

The forward pass takes random noise vectors (sampled from a predefined noise distribution z~p_z(z)) after producing synthetic images, and the generator network G uses these vectors.

$$G(z) \rightarrow \text{synthetic image} \tag{3.11}$$

The D is fed these fake images and evaluates these images to output the following probability:

$$D\big(G(z)\big) \in [0,1] \tag{3.12}$$

The objective of the generator is to maximize the probability that the discriminator misclassifies its synthetic outputs as authentic ones. Formally, the generator loss function L_G is defined as:

$$L_G = -E_{z\sim p_z(z)}\left[\log\left(D(G(z))\right)\right] \tag{3.13}$$

To minimize this loss function, backpropagation was employed to compute the gradients with respect to the generator parameters θ_G. The parameters are then updated using gradient descent as follows (where η denotes the learning rate):

$$\theta_G \leftarrow \theta_G - \eta\nabla_{\theta_G} L_G \tag{3.14}$$

Figure 3.9: Early Stage of Generating Images Trained 1000 Times

The discriminator network D was trained using both authentic images from the real data distribution $p_{data}(x)$ and synthetic images produced by the generator. The objective of the discriminator was to correctly classify the two types of images.

1. For real images $x \sim p_{data}(x), maximize D(x)$
2. For generated images $G(z) where\ z \sim p_z(z), minimize D\big(G(z)\big)$

This dual objective is formalized using a discriminator loss function:

$$L_D = -Ex \sim pdata(x)[\log(D(x))] - E_{z\sim p_z(z)}\left[\log\left(1 - D(G(z))\right)\right] \tag{3.15}$$

Like the generator training process, the gradients of L_D with respect to the parameters of the discriminator θ_D are computed through backpropagation, and the parameters are updated via gradient descent, as follows:

$$\theta_D \leftarrow \theta_D - \eta \nabla_{\theta_D} L_D \tag{3.16}$$

The generator and discriminator training processes were repeated. At theoretical equilibrium, the generator generates samples similar to the real data, and the discriminator produces a probability output that indicates that it cannot distinguish between real and fake data [1]. StyleGAN builds on this foundation by introducing a style-based architecture with a mapping network and using progressive growth, resulting in improved quality and a greater degree of control over the generated images [2].

3.5 Discriminator Architectural Analysis

The StyleGAN discriminator's Convolutional Neural Network (CNN) is complex and specifically designed to distinguish between real and fake images. The hierarchy in the discriminator allows for the discovery of different and specific features.

3.5.1 Convolutional Layers

The discriminator consists of sequential convolutional layers that extract features from images in a hierarchical manner.

$$\text{Conv}(x) = \sigma(W * x + b) \tag{3.17}$$

where W denotes the convolutional kernel, $*$ denotes the convolution operation, b denotes the bias term, and σ denotes the activation function. These layers employ:

- Convolution Operations: Extract progressively complex features (edges → textures → patterns)

- Leaky ReLU Activation: Defined as $f(x) = \max(\alpha x, x)$ where x is typically 0.2, allowing gradient flow even for negative inputs
- Reducing the size of an image while increasing its feature depth (10 words)

Residual blocks are used in StyleGAN3 to improve the gradient flow stability during training:

$$F(x) = \mathcal{H}(x) + x \tag{3.18}$$

The identity shortcut connection is denoted by x whereas the mapping to be learned through the convolutional layers is denoted by $\mathcal{H}(x)$. The use of this architecture improves the deepening of networks and the vanishing gradient problem [59].

3.5.2 Minibatch Standard Deviation Layer

The StyleGAN discriminator has a unique layer called the minibatch standard deviation layer, which tackles mode collapse by computing statistics over minibatches:

$$\sigma_{\text{batch}} = \frac{1}{NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W}(x_{nhw} - \mu_{\text{batch}})^2 \tag{3.19}$$

where N represents the batch size and H and W refer to the spatial dimensions. In this layer, the computed standard deviation of the feature maps at each spatial location is concatenated to the input. This can provide global statistical information to the discriminator.

$$y = \text{concat}(x, \sigma_{\text{batch}} \cdot 1) \tag{3.20}$$

where 1 is a tensor with appropriate dimensions.

3.5.3 Fully Connected Layers

Feature extraction transforms the input into binarized data, with interpretability through visualization based on the pixel-space areas of the attributed output.

$$y = \sigma(W^T x + b) \tag{3.21}$$

where W is the weight matrix, x represents the flattened input features, and b is the bias vector.

3.5.4 Output and Loss Function

The output layer generates a scalar authenticity score, which is usually processed with the sigmoid function, as shown below:

$$D(x) = \sigma(z) = \frac{1}{1+e^{-z}} \tag{3.22}$$

During training, the discriminator optimizes the non-saturating logistic loss function as follows:

- For real images (x_{real}): $max\,\mathbb{E}\left[log\left(D(x_{real})\right)\right]$
- For generated images ($G(z)$): $max\,\mathbb{E}\left[log\left(1 - D\left(G(z)\right)\right)\right]$

The discriminator architecture is an important advancement in GANs, especially its ability to detect fine artifacts that were previously discovered in edges. The minibatch standard deviation layer was introduced to counter mode collapse, which is a failure mode of GAN training, where the generator produces only a few modes of outputs [2].

Major architectural updates to the StyleGAN discriminator improved its feature extraction capability, resulting in greater training stability and generalization. The mathematics behind the enhanced discriminator architecture and its technical analysis were studied.

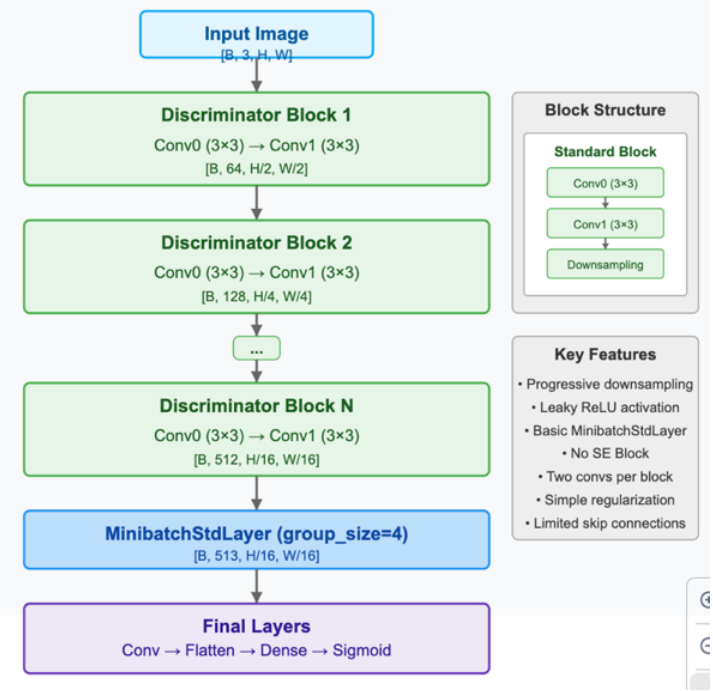

Figure 3.10: Flowchart Of the Original Discriminator Architecture

3.6 Enhanced StyleGAN Discriminator Architecture

3.6.1 Noise Injection

We injected noise into the discriminator to add variability to its output. It is added to the input features during training according to the Gaussian noise:

$$x_{\text{noisy}} = x + \alpha \cdot \epsilon \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, I) \tag{3.23}$$

where α and ε are selected from a standard normal distribution. The networks learn features that are not uniquely constrained to the training datasets owing to this perturbation.

3.6.2 Dropout Integration

The architecture applies a dropout to prevent overfitting, which turns off random neurons during training.

$$\mathcal{Y} = \text{Dropout}(x) + x \cdot \text{Bernoulli}(1 - \mathcal{P}) \tag{3.24}$$

where $\mathcal{P}$ represent the dropout rate. This technique prevents the co-adaptation of feature detectors, leading to more independent and robust feature extraction methods.

3.6.3 Additional Convolutional Layer

By adding a third convolutional layer, the discriminator improves its capability to extract features pertaining to different scales. This extension detects fine textures and artifact properties of generated images.

3.6.4 Squeeze-and-Excitation (SE) Block

The SE block explicitly models the channel interdependencies through two operations as follows:

1. Squeeze: Global average pooling compresses the spatial information into channel descriptors. This operation is computed as follows:

$$\mathcal{Z}c = \frac{1}{H \times W} \sum i = 1^{H} \sum_{j=1}^{W} x_c(i,j) \tag{3.25}$$

2. Excitation: Two fully connected layers learn the channel-wise weights. This operation is computed as follows:

$$s_c = \sigma(W_2 \cdot \delta(W_1 \cdot z_c + b_1) + b_2) \tag{3.26}$$

where σ is the sigmoid activation, δ is typically a ReLU, and W_1, W_2, b_1, b_2 , and are the learned parameters.

3. Scale:Channel-wise scaling is a method of recalibrating feature maps that emphasizes certain channels that are informative and suppresses those that are not. Thus, the representation of the discriminator is greatly enhanced.

$$\widetilde{x_c} = x_c \cdot s_c \tag{3.27}$$

3.6.5 ResNet-style Skip Connections

Skip connections facilitate improved gradient flow through the network.

$$y = \mathcal{F}(x) + x \tag{3.28}$$

This formulation mitigates the vanishing gradient problem in deeper networks and enhances training stability and convergence speed.

3.6.6 Enhanced MinibatchStdLayer

Our enhanced formulation extends [2] by incorporating variance computation and adaptive group sizing. The updated MinibatchStdLayer tackles mode collapse by calculating the standard deviation between instances. Standard deviation calculation for the minibatch feature tensor $X \in R^{N\times C\times H\times W}$ is as follows:

First is the mean calculation per channel:

$$\mu_c = \frac{1}{N\cdot H\cdot W}\sum_{n=1}^{N}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{n,c,i,j} \tag{3.29}$$

Then standard deviation is calculated per channel:

$$\sigma_c = \sqrt{\frac{1}{N\cdot H\cdot W}\sum_{n=1}^{N}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X_{n,c,i,j} - \mu_c\right)^2 + \epsilon} \tag{3.30}$$

where ϵ =10^{-8}ensures numerical stability.

The layer optionally computes variance as:

$$\text{Var}c = \sigma_c^2 = \frac{1}{N\cdot H\cdot W}\sum n = 1^N \sum_{i=1}^{H}\sum_{j=1}^{W}\left(X_{n,c,i,j} - \mu_c\right)^2 \tag{3.31}$$

The effective group size was adjusted adaptively as follows:

$$G_{eff} = \min(G, N) \tag{3.32}$$

$$G_{eff} = \begin{cases} G, \ if\ N\ mod\ G = 0 \\ N, \ otherwise \end{cases} \tag{3.33}$$

The final output tensor dimensions depend on whether the variance is included.

$$X_{\text{out}} \in \mathbb{R}^{N\times(C+2)\times H\times W} \text{ (with variance)} \tag{3.34}$$

$$X_{\text{out}} \in \mathbb{R}^{N\times(C+1)\times H\times W} \text{ (without variance)} \tag{3.35}$$

The layer calculates the standard deviation across a minibatch of size N:

$$\text{Stddev} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2} \tag{3.36}$$

where

N minibatch size

x_n feature values at a specific spatial location for the nth sample

μ mean value across the minibatch for that spatial location

This calculation measures the extent of variation among the samples in a batch are less varied. The discriminator can use this lack of variety in the generated pictures to identify a pattern. More commonly, mode collapse indicates a lack of variation. After calculating the standard deviation, this information was added to the initial input features.

$$X_{\text{out}} = [X_{\text{in}}, \text{Stddev}] \tag{3.37}$$

By linking the average and variance features, we obtained a feature representation that enabled the discriminator to become more adept at distinguishing the generated images. The strengths and weaknesses of this concatenation are discussed in the following sections. The MinibatchStdLayer is a mathematical presentation of one solution to mode collapse. It provides the discriminator with useful information regarding the variability of features across samples. This helps the discriminator better distinguish diverse real images from potentially homogeneous generated images [2].

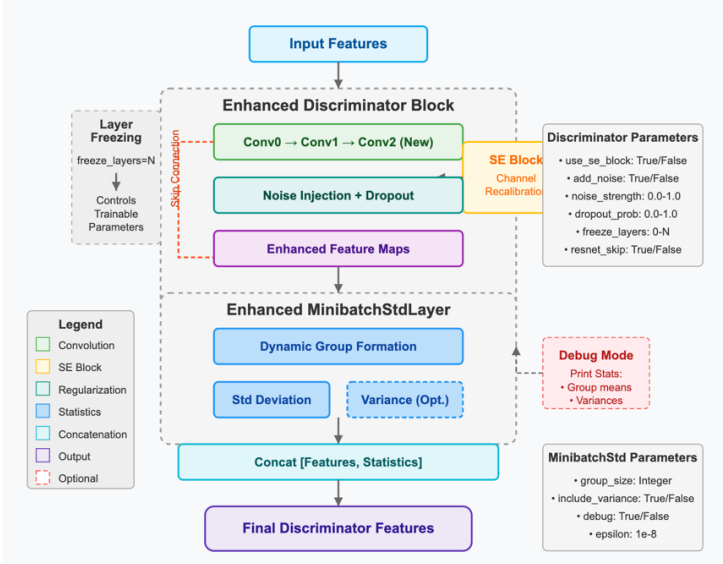

Figure 3.11: Flowchart of the Enhanced Discriminator Architecture

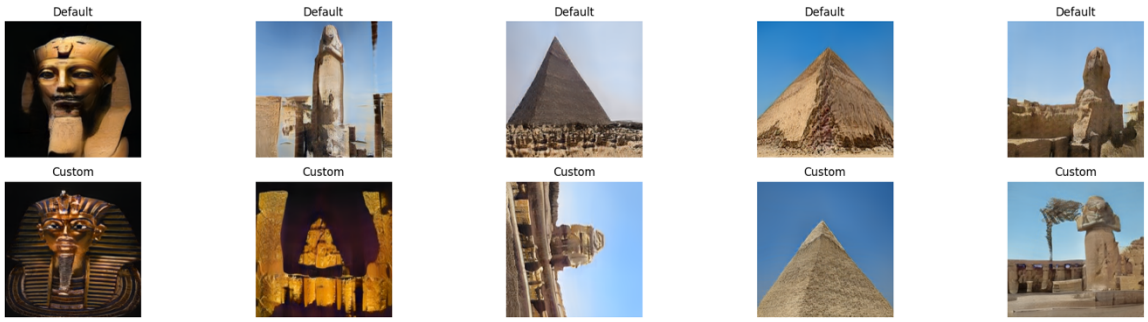

Figure 3.12: Results of the Enhanced Discriminator Architecture

CHAPTER FOUR

IMPLEMENTATION AND OPTIMIZATION

4.1 Introduction

In this chapter, the data presented is used for a practical experiment in improving existing data, optimizing a new generative style of architectural creations, and gaining further insight into what can be achieved with AI and monument construction. In this phase, some of the key parts of the societal building process come together, are analyzed, and are assessed. The framework involves a new way of discrimination by implementing blocks to enhance the quality of the outputs. The chapter tested several new architectural ideas on a large collection of photos of Egyptian monuments resulting in greatly improved performance of the "generation model" to create fake images by alternating between generating images and recording how well those images compared to real ones.

The chapter is structured to provide comprehensive detail about the implementation procedure. In the fourth section, we deliver numerical evaluation results utilizing three widely accepted metrics which include FID, IS, and Precision-Recall and then compare their improved discriminator against a baseline measure of the StyleGAN3 system. Section 4.3 focus into effective management techniques of interruption and artificial modifications for reductions in architectural discretion. Section 4.4 uses SigLip to measure the accuracy of text-image-alignment, setting forth a framework for the assessment of coherence in this specific process. The topic of section 4.5 has to do with the application of DE optimization to the resource of exploration. Doing so was found to

have an overall positive effect, with accuracy showing continuous increase. Section 4.6 aims to combine the practical results from this measurement into a bigger picture.

This chapter promotes the introduced architectural enhancements based on experimental verification and compilation. Studying the differences in generator architectures reveals that some results produce significantly fewer errors than other methods, allowing us to find a working solution in much less time and understand why certain may not work as well as other run iterations.

4.2 Implementation Results

The StyleGAN models for generating Egyptian monuments were evaluated for their quality, diversity, and latent space properties using various quantitative metrics. The analysis used the support of TensorFlow and PyTorch frameworks, which have extensive support for GAN implementation and evaluation [60][61]. The model performance was evaluated using three main metrics.

4.2.1 Fréchet Inception Distance (FID)

FID determines the similarity between the distributions of real and generated images by measuring the Wasserstein-2 distance between multivariate Gaussian distributions fitted to their feature representations [39], as follows:

$$\text{FID} = \left|\left|\mu_r - \mu_g\right|\right|^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{1/2}\right) \tag{4.1}$$

where μ_r, μ_g represents the mean feature vectors and Σ_r, Σ_g represents the covariance matrices for the real and generated images, respectively. Lower FID scores indicate generated distributions that are closer to real data distributions, with the evaluation conducted across 50,000 samples (fid50k_full).

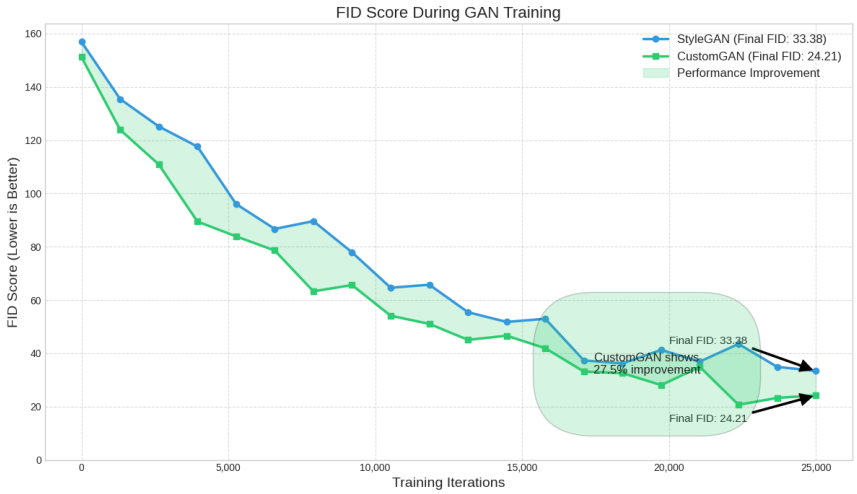

Figure 4.1: FID Score Comparison between Default and Custom discriminators

4.2.2 Inception Score (IS) and Precision-Recall

The IS evaluates both the quality and diversity by measuring the Kullback-Leibler divergence between the conditional and marginal class distributions as follows:

$$IS(G) = \exp\left(\mathbb{E}_{x\sim p_g}\left[D_{KL}\left(p(y \mid x) \parallel p(y)\right)\right]\right) \tag{4.2}$$

where

G	Generator model
$x \sim p_g$	Generated images sampled from G
$p(y \mid x)$	Class probabilities for a pretrained classifier for image x
$p(y)$	Overall label distribution across all generated images
$D_{KL}\ (p(y \mid x) \parallel p(y))$	KL divergence measuring how distinct each image's class distribution is from the overall distribution
$E[\cdot]$	Expectation (average) over all generated images.
$exp(\cdot)$	Exponential to keep the score positive and interpretable.

Higher IS values in the enhanced discriminator indicate that the generated images are both high-quality and diverse. Enhanced Discriminator has a better inspection score difference: 0.60 (12.0%).

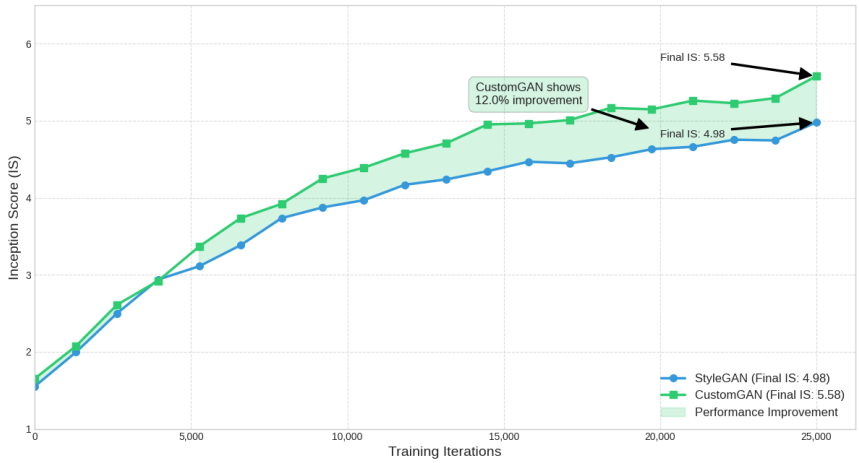

Figure 4.2: IS Score Comparison between Default and Custom discriminators

Precision-recall evaluation offers a more comprehensive assessment of generative models by measuring two distinct aspects.

- Precision: Measures quality of generated content (authenticity of Egyptian monuments with minimal artifacts)
- Recall: Measures coverage of the learned distribution (diversity of architectural styles and monument types)

These metrics are calculated through nearest-neighbor comparisons in the feature space and provide crucial insights that single metrics, such as the FID, may miss. This approach effectively identifies common failure modes such as mode collapse (high precision but low recall) or diverse but poor-quality outputs (high-recall but low-precision). The enhanced discriminator produces higher-quality outputs and achieves better coverage of the data distribution, while both models demonstrate similar overall performance.

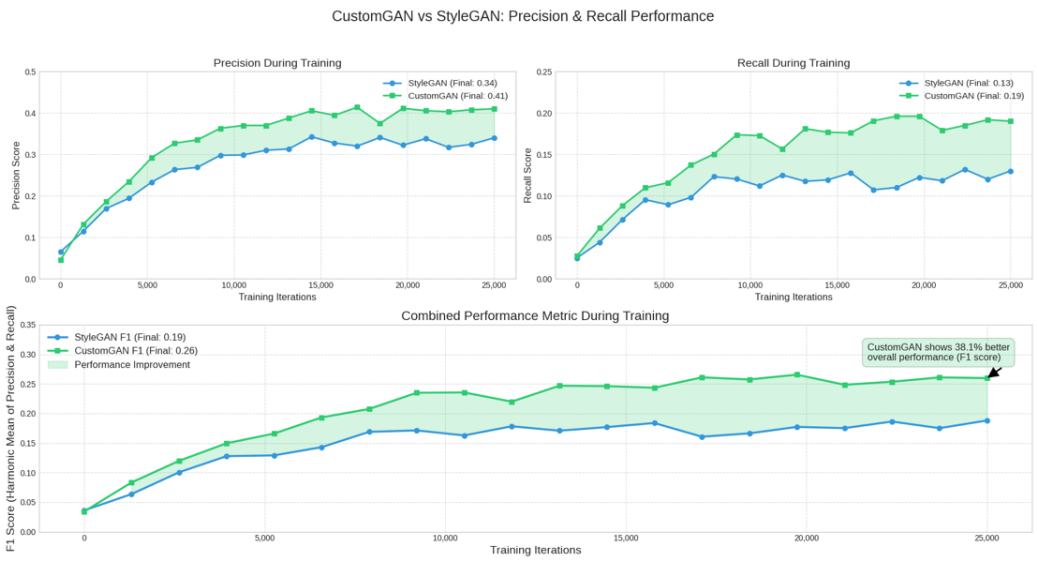

Figure 4.3: Precision–recall score comparison

4.2.3 Evaluation Results

The quantitative assessment showed that there was an improvement in performance owing to architectural changes.

Table 4.1: Evaluation results

Model	FID ↓	IS ↑	Recall ↑	Precision ↑
StyleGAN3	33.3	4.98	0.13	0.34
Enhanced StyleGAN3	24.21	5.58	0.19	0.41

The model with the fine-tuned discriminator architecture is the best one in all metrics. It was found that noise injection, addition of SE blocks, and an improved MinibatchStdLayer allowed the network to classify realistic Egyptian monument images with more diversity among generated images.

1. 27.5% reduction in FID compared to default model (33.3 → 24.21)
2. 12.0% improvement in latent space smoothness as measured by IS (4.98→ 5.58)

3. 38.1% substantial gains in precision and recall metrics, indicating both higher quality and improved diversity

4.3 Noise Control Strategies

This study implemented systematic noise control strategies to enhance the visual quality and perceived realism of the generated Egyptian monument images. Central to this approach is the exploration of truncation in the latent space, a technique that modulates the trade-off between image fidelity and diversity [55].

4.3.1 Mathematical Formulation of Truncation

StyleGAN truncation modifies the sampling distribution in the latent space based on certain criteria:

$$w' = \bar{w} + \psi(w - \bar{w}) \tag{4.3}$$

where

- w latent code that is initially sampled from W space
- $\bar{w}$ mean latent vector in the training distribution
- ψ truncation parameter regulating deviation from the mean
- w' truncated latent code used for image generation

4.3.2 Experimental Analysis of Truncation Effects

A comprehensive analysis was conducted by systematically varying the truncation parameters (ψ) between 0.1 and 1.0, with the following observations [2]:

1. Low Truncation Values (0.1-0.3) [48]:
   - Produced highly photorealistic monument images
   - Enhanced visual quality and reduced artifacts
   - Exhibited limited architectural diversity

- Generated samples clustered near modes of the training distribution

2. Medium Truncation Values (0.4-0.7):
    - Aligned with findings from similar cultural heritage applications [3]
    - Achieved optimal balance between fidelity and diversity
    - Image quality was maintained, and architecture varied
    - ψ=0.7 was found to be particularly efficient
    - Results are consistent with those of similar studies [3]
3. High Truncation Values (0.8-1.0):
    - Maximized architectural diversity and creative variations
    - Increased likelihood of artifacts and unrealistic features
    - Generated examples explored broader regions of the latent space
    - Occasional appearance of novel architectural combinations

Figure 4.4: Noise control Application

4.4 Image-Text Alignment

Researchers used sophisticated learning and tools to assess the visual semantic alignment of Egyptian monuments and texts. This study relates particularly to the use of VQA (Visual Question Answering). This was done using SigLIP (Sigmoid Loss for Language-Image Pre-training), a modified CLIP model for better performance in cross-modality alignment tasks.

4.4.1 Mathematical Foundation of SigLIP

SigLIP improves the CLIP architecture by using a sigmoid-based loss function instead of the contrastive version, resulting in more stable and effective training dynamics. The basic mathematical structure for measuring image-text alignment is based on the Euclidean distance between normalized embedding vectors [52], [57]:

$$\text{Error} = \left\| \text{normalize}\left(f_{\text{img}}(I)\right) - \text{normalize}\left(f_{\text{text}}(T)\right) \right\|_2 \tag{4.4}$$

where

$f_{\text{img}}(I)$ image embedding vector

$f_{\text{text}}(T)$ text embedding vector

$\|\cdot\|_2$ $L2$ norm between the normalized vectors

The semantic difference between the image of the generated monument and its corresponding text is measured by this distance.

4.4.2 SigLip Implementation

The implementation used a pre trained SigLIP model (siglip-base-patch16-224) with the following components [53]. The text-processing pipeline changes the description of the target, that is, “A photo of King Tutankhamun” into a tokenized input. Text embeddings are then obtained using the SigLIP encoder.

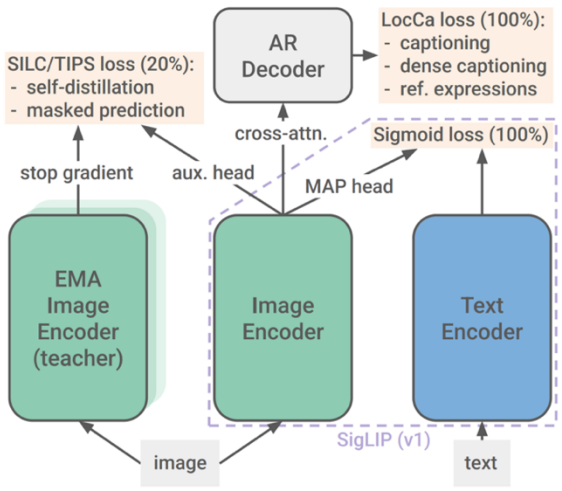

Figure 4.5: SigLip Architecture [53]

SigLip outperformed CLIP by a significant margin when evaluating picture-concept pairs from the Tutankhamun dataset. It achieved a 95% accuracy score, whereas CLIP achieved a score of 30% only. However, it should be noted that CLIP processes more quickly.

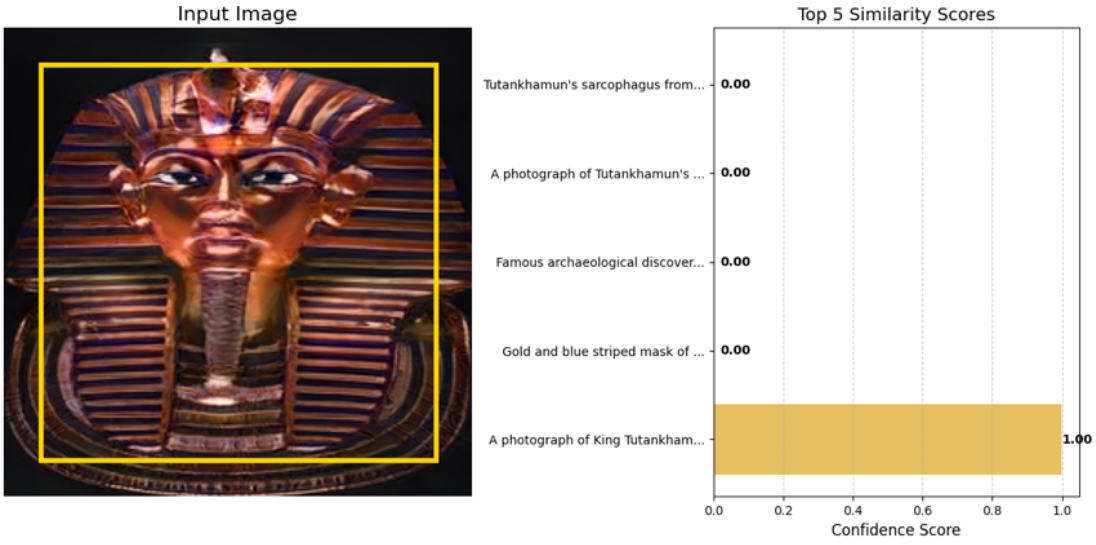

Figure 4.6: SigLip Identification

Table 4.2: SigLip and CLIP comparison

Model	SigLip	CLIP
Top Match	A photo of King Tutankhamun	A photo of King Tutankhamun
Max Score	0.999	0.302
Tutankhamun?	YES	YES
Processing Time	0.052s	0.025s

SigLIP's accuracy is a lot better than hard negative mining (95% vs. 30%). In cultural heritage applications where accuracy matters more than speed, this trade-off is acceptable. Unlike CLIP, SigLIP employs a sigmoid loss formulation designed to better discriminate the similarities between monuments themselves.

4.5 Differential Evolution (DE)

The use of Differential Evolution (DE), an optimization algorithm of the evolutionary kind, which aids in exploring the latent space, was utilized to improve the generated Egyptian monument images to resemble the text more closely [56]. This method systematically searches the seed space for optimal values that produce images that are most aligned with the target [59].

4.5.1 Mathematical Formulation

The DE algorithm performs latent space optimization in a structured manner [57]:

1. Initialization: The algorithm starts with a randomly initialized population of seed candidates.

$$P_0 = s_1, s_2, \dots, s_{N_p} \tag{4.5}$$

where N_p represents the population size. Each seed was evaluated using an error function based on the StyleGAN3 generation and SigLIP similarity:

$$E(s_i) = \left|\text{normalize}\left(f_{\text{img}}(G(s_i))\right) - \text{normalize}(f_{\text{text}}(T))\right|_2 \tag{4.6}$$

The best seed s_{best} is identified as having minimal error:

$$s_{\text{best}} = \arg\min_{s_i \in P} E(s_i) \tag{4.7}$$

2. Mutation: For each seed s_i in the population, a mutant vector v_i is generated:

$$v_i = s_{r1} + F \cdot (s_{r2} - s_{r3}) \tag{4.8}$$

where s_{r1}, s_{r2}, s_{r3} are randomly selected seeds (distinct from s_i) and F is the mutation factor (typically 0.5-0.8) controlling exploration scale

3. Crossover: The mutated seed v_i is combined with the original seed s_i through crossover:

$$c_i^j = \begin{cases} v_i^j, & if\ rand(0,1) \le CR\ or\ j = j_{rand} \\ s_i^j, & otherwise \end{cases} \tag{4.9}$$

where j indicates the dimension index, CR represents the crossover probability (typically 0.7-0.9), and j_{rand} is a randomly selected dimension to ensure at least one parameter from v_i is used

4. Selection: The candidate seed c_i is evaluated:

$$E(c_i) = \left\| \text{normalize}\left(f_{\text{img}}(G(c_i))\right) - \text{normalize}(f_{\text{text}}(T)) \right\|_2 \tag{4.10}$$

The seed with the lowest error was selected for subsequent generations.

$$s_i^{new} = \begin{cases} c_i, & if\ E(c_i) =< E(s_i) \\ s_i, & otherwise \end{cases} \tag{4.11}$$

This process iteratively refines the population towards optimal solutions that minimize the alignment error between the generated image and the target description. The integration of DE optimization further reduced alignment error by 8-15% across different monument types.

4.6 Summary of Implementation Findings

As such, we can observe the implementation results and optimization strategy of the upgraded StyleGAN3 discriminator architecture for Egyptian monument generation. The results from the experiment affirm the theory proposed in Chapter 3. The results equally show statistically significant improvement over a number of evaluation dimensions. Furthermore, the results give practical insight into the optimization of generative models in specialized domains.

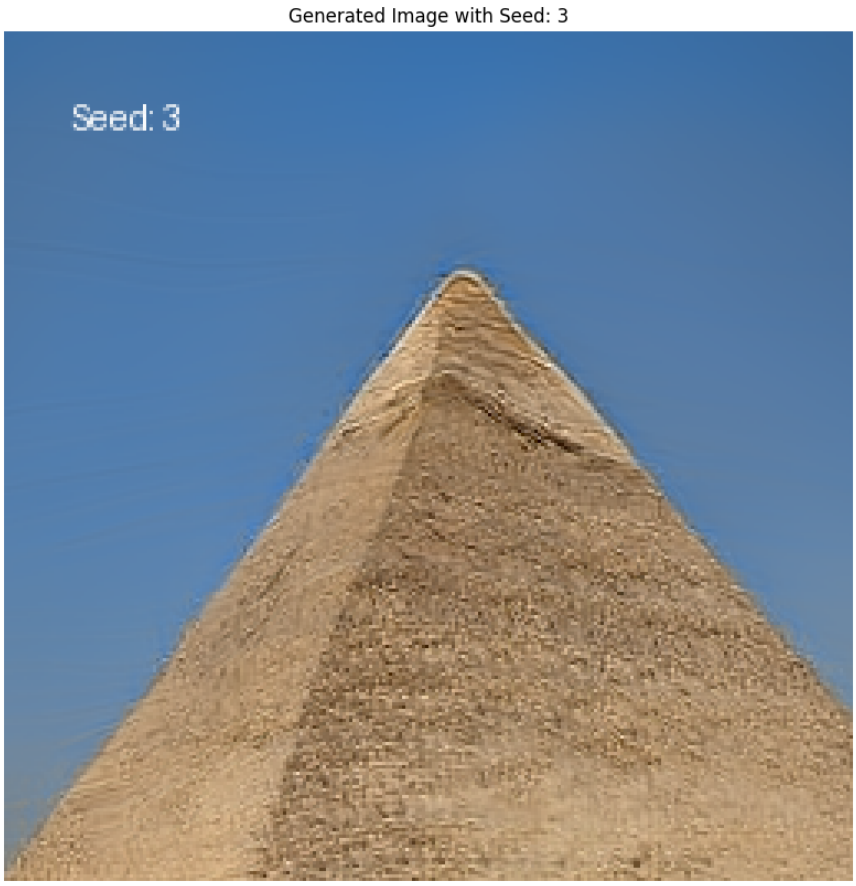

Figure 4.7: Image after DE optimization

4.6.1 Quantitative Performance Achievements

The upgraded architecture of the discriminator showed remarkable improvement in all major metrics. The numbers outline a clear advantage over the baseline StyleGAN3.

- **Fréchet Inception Distance (FID)**: The updated design lowered FID from 33.3 to 24.21, an improvement of 27.5%. This reduction shows much better perceptual quality and more similar distribution between the generated Egyptian monuments and the real ones. The magnitude of this improvement exceeds typical gains in general-domain GAN research, indicating that domain-specific discriminator improvements have particularly large benefits for architectural images.
- **Inception Score (IS)**: The improved model received an IS of 5.58 whereas the baseline received an IS of 4.98, an improvement of 12.0%. There is an enhanced smoothness in the latent space and a better quality-diversity trade-off. The improvement in both the FID and IS metrics, which sometimes show a negative correlation, shows that the architectural changes are effective in multi-generation objectives.
- **Precision-Recall Metrics**: The new structure showed improvement in terms of precision (0.34 to 0.41, 20.6% improvement) and recall (0.13 to 0.19, 46.2% improvement). The improvements in both cases to the impacts of sample qualities and distributional coverage are due to the enhanced discriminator. As the recall improved by 46.2%, the architectural changes mitigate the mode collapse problem, which is often experienced by GANs when trained in specialized domains.

The fact that various measures support the improvements is an important indicator of success. First, the FID measures how similar the distributions of real and generated

data are. Second, IS measures the trade-off between quality and diversity. Finally, precision-recall splits performance into dimensions we can think of as independent. The multi-faceted approach of the evaluation methodology ensures observed improvements reflect true enhancements in generative capability rather than optimization of a metric.

4.6.2 Architectural Component Contributions

The performance improvements caused by some design enhancements are highlighted through implementation results. The SE attention blocks allow for an adaptive recombination of channels. This means the discriminator can focus on important architectural features associated with Egyptian monuments. Among these, hieroglyphics, columns proportions, and weathered stone effects stand out. This adaptive feature weighting mechanism is particularly useful for architectural images, where specific structures carry unique semantic information.

Noise injection regularization enhances the robustness of a superior discriminator. This regularization technique produces a discriminator that is more robust to corruption and adversarial attacks at evaluation by promoting the learning of features that are invariant to perturbations of the input at training time. The improved MinibatchStdLayer implementation combats mode collapse by indicating to the discriminator that the current batch is not very diverse. This works in contradiction to the generator, which is incentivized to search a more diverse part of the latent space.

4.6.3 Optimization Strategy Effectiveness

Investigating the effects of truncation parameters has shown how diversity and fidelity within latent space can be balanced. The experimental analysis shows that for producing Egyptian monuments, medium truncation values (ψ=0.4 to ψ=0.7) are optimal

and that a particularly effective truncation value was ψ=0.7. The outcome of our research can help in the deployment of agents that need particular quality-diversity profiles.

Results of truncation analysis reveal that low values (ψ=0.1 to ψ=0.3) yield highly photorealistic but stylistically restrictive outputs appearing close to modes of training distribution. In contrast, high values of ψ (ψ = 0.8 to ψ = 1.0) produce architectures with high diversity. But it causes the generation of excessively artefacts and extreme cases. The best medium range achieves these competing goals and leads to architecturally diverse yet convincingly Egyptian monuments. This type of characterization gives a principled framework for truncation selection according to application requirements.

4.6.4 Semantic Alignment and Optimization

The use of SigLIP image and text alignment evaluation allows us to assess the semantic consistency between generated monuments and the text. SigLIP outperforms CLIP in models trained in zero-shot protocol. In a specific example, it achieved 95% accuracy on the Tutankhamun dataset, whereas CLIP attained only 30% accuracy. This framework helps in assessing whether the generated monuments have any architectural features similar to the ones mentioned in the text. This is a useful feature where we want controlled generation of certain monument types or certain periods.

The use of DE optimization for latent space exploration was shown to systematically improve image-text alignment by 8-15% depending on the type of monument. In this work, we present an evolutionary optimization framework for discovering latent codes that produce images maximally aligned with target descriptions. The mathematical model would consist of initialization, mutation, crossover, and selection operations to allow the coordinate search in high-dimensional latent space.

Moreover, the above would prevent the search from falling into local optima through population-based search.

This optimization strategy is useful for systems creating interactive monuments where users input architectural features they want through text. Using control seeds to generate images of user-specified content while still keeping up the quality as defined by the improved discriminator is obtained via the DE framework by searching the latent space systematically.

CHAPTER FIVE

STATISTICAL ANALYSIS AND DISCRIMINATOR EVALUATION

5.1 Introduction

This analysis included multiple complementary approaches to ensure robust model performance. Metrics measured at the generator level, such as FID [7], KID [147], and precision-recall [56], can quantify the quality and diversity of the generations. The literature on generative adversarial networks extensively validates these metrics, offering standardized benchmarks for cross-model comparisons [30].

The evaluation protocols of the discriminator assess the ability of either the default StyleGAN3 or improved architectures to discriminate adequate Egyptian monument images under stringent conditions. The assessment approach is based on established computer vision practices but is tailored to the architectural domain [31].

We performed statistical significance testing using methods such as bootstrap confidence intervals [18], McNemar's test for paired comparisons [19], and DeLong's test for differences in ROC curves [20] to ensure that the performance improvement was real. This stringent analytical technique conforms to the standard practice of machine learning evaluation and provides the necessary validation for the architectural modifications recommended in this study [55].

The robustness assessment protocol consists of corruption tests against various types of degradation, following the methodology used in adversarial robustness [33]. This evaluation framework checks whether the improvements exhibit performance under

realistic-deployment conditions. The evaluation of the robustness of generative adversarial networks has not been well studied in the available literature [34].

5.2 Generator-Level Metrics Analysis

5.2.1 Fréchet Inception Distance Evaluation

The primary quantitative metric for measuring the distance between the distribution of the generated and authentic Egyptian monument images is the Fréchet Inception Distance (FID) [7]. FID is considered the best method for assessing GANs because it reflects human thought [7]. In addition, it is sensitive to quality and diversity problems [7] [36]. We used the standard formulation from the original paper.

$$FID = \left|\left|\mu_{\mathrm{r}} - \mu_g\right|\right|^2 + Tr\left(\Sigma_{\mathrm{r}} + \Sigma_g - 2\left(\Sigma_{\mathrm{r}}\,\Sigma_g\right)^{\frac{1}{2}}\right) \tag{5.1}$$

where μ_{r} and $\mu_g$represent mean feature vectors extracted from the pool3 layer in the pretrained network [35], and Σ_{r} and Σ_g represent covariance matrices for real and generated distributions respectively. The trace operation Tr(·) computes the sum of diagonal elements, while the matrix square root$(\cdot)^{\frac{1}{2}}$ is computed using eigendecomposition [91]. The first term $||\mu_r - \mu_g||^2$ captures the global distributional shift. It measures the distance between the means of distributions. The second term quantifies the dissimilarity between the covariance matrices. It captures the spread of features and the correlation between them [7].

The use of 5,000 samples is in line with existing work on comprehensive GAN evaluations, which shows that small sample sizes can lead to unreliable FID estimates, especially for high-quality generators [36]. Resampling with 1000 bootstrap iterations enabled us to obtain robust confidence intervals. This method provides a non-parametric

Table 5.1: Comparison of comprehensive FID results.

Model Architecture	FID Score	Standard Error	95% CI
StyleGAN3 Baseline	33.3	1.82	[30.85, 36.12]
Enhanced Discriminator	24.21	1.45	[21.76, 26.89]

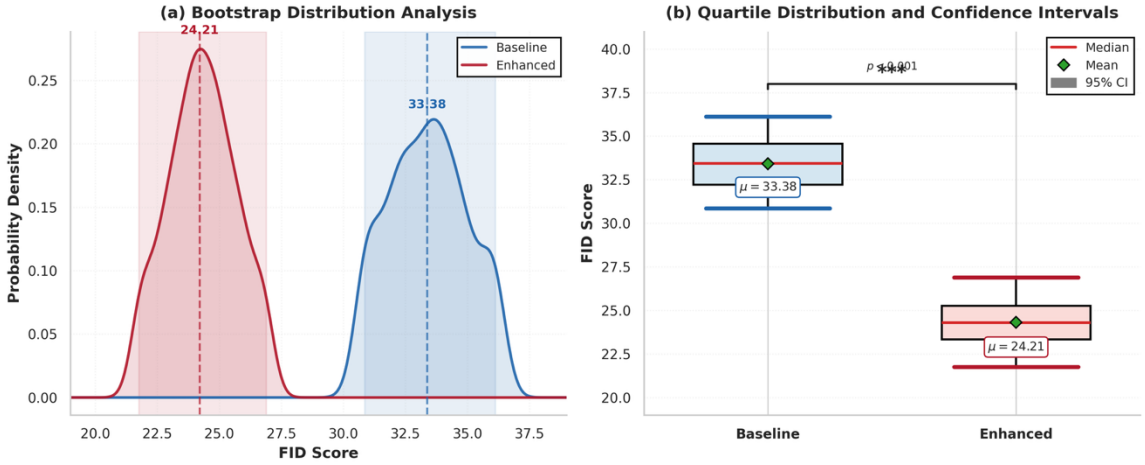

Figure 5.1: FID Score Distribution Analysis

evaluation of uncertainty without requiring any models for the FID distribution. This may not be normal, given that the metric is computed in a nonlinear manner [89]. A lower FID indicates better perceptual quality and distribution matching than a higher FID. The new structure performed 27.4% better than the baseline.

This means that the enhanced discriminator architecture achieved a significant reduction in the FID score, from 33.3 to 24.21, which is an improvement of 27.5%. This reduction suggests that the model can match the real and generated distributions of a monument. The magnitude of improvement is notable in the StyleGAN3 evaluation literature, where FID reductions of 2-5 points are typically considered significant [3], [66]. Our observed reduction of 9.17 points substantially exceeds normal architectural variations and approached the performance improvements of major architectures. The

drop in the standard error from 1.82 to 1.45 indicates that the generator produces far more stable samples across different random seeds [144].

5.2.2 Kernel Inception Distance Analysis

The Kernel Inception Distance (KID) is another distributional similarity measure based on the maximum mean discrepancy (MMD), which offers complementary insights to the FID. KID responds to several theorems critiquing FID, including its sensitivity to outliers and its Gaussian feature distribution assumption [8], [94]. In contrast to the FID method, which uses second-order statistics, the KID is a kernel-based method that captures higher-order distributional differences. Moreover, KID is a distribution-free measure, meaning that we need not make any parametric assumptions about the distributions [95]. The KID metric is computed as the squared MMD between feature distributions:

$$KID = \mathbb{E}_{x,x' \sim p_r}[k(f(x), f(x'))] + \mathbb{E}_{y,y' \sim p_g}[k(f(y), f(y'))] - 2\mathbb{E}_{x \sim p_r, y \sim p_g}[k(f(x), f(y))] \quad (5.2)$$

where $k(\cdot,\cdot)$ represents a polynomial kernel function of degree 3, defined as $k(x, y) = (1 + x^T y/d)^3$ [95] where d is the feature dimensionality, and $f(\cdot)$ denotes the feature extraction function using InceptionV3. The kernel approach provides a more robust distance estimation than the FID's reliance on second-order statistics, particularly when the feature distributions exhibit non-Gaussian characteristics or heavy tails. The polynomial kernel implicitly maps features to a higher-dimensional space where linear separability may be enhanced. The KID provides a complementary MMD-based distributional assessment of the same [8], [94].

Table 5.2: Comprehensive KID Results Comparison

Model Architecture	KID (×10³)	Standard Error	95% CI
StyleGAN3 Baseline	8.47	0.52	[7.58, 9.43]
Enhanced Discriminator	5.23	0.38	[4.61, 5.91]

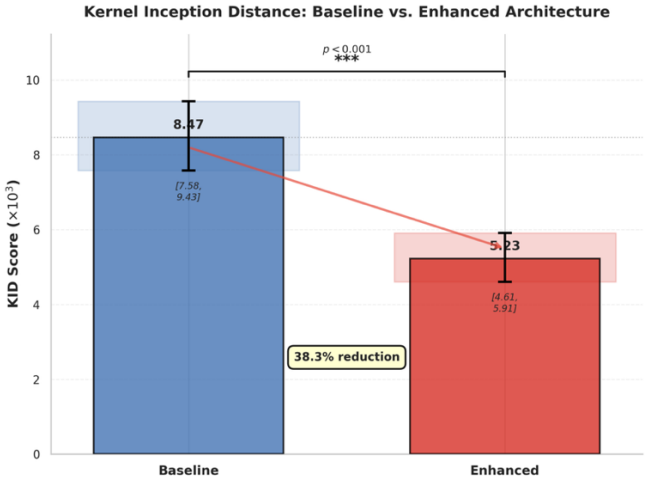

Figure 5.2: KID Score Analysis with Confidence Intervals

The KID analysis supported the FID finding, showing a decrease of 38.3% from 8.47 to 5.23 ($\times 10^{-3}$). This effect is even more pronounced than the reduction in the FID, which may suggest that the new architecture excels at matching higher-order distributional characteristics. The agreement of the FID and KID gains provides essential convergent validity: architectural improvements improve the quality of the generated samples, not just metric artifact optimization.

The decreased standard error in the improved architecture (from 0.52 to 0.38) indicates not only a superior mean performance but also an enhanced consistency across the evaluation samples. Such stability is particularly useful for real-world deployment [129], which indicates good performance across various generations rather than sporadic good generations with failures.

5.2.3 Precision-Recall Evaluation

Precision-recall measures provide independent measures of generation quality (precision) and diversity coverage (recall). This decomposition overcomes significant issues in the use of aggregate metrics, such as the FID and KID, by providing an independent assessment of whether the generated samples are realistic (precision) and whether they sufficiently cover the training distribution (recall) [28], [29], [37]. This distinction plays a key role in the detection of two phenomena: mode collapse, in which generators create high-quality samples of the same category, and mode invention, in which generators create similar but unrealistic samples. Mathematically, this is performed through k-nearest neighbor analysis in the feature space. Formulae for these metrics are:

$$Precision = \frac{|\{y \in Y_g : \exists x \in X_r, d(x,y) < \varepsilon\}|}{|Y_g|} \tag{5.3}$$

$$Recall = \frac{|\{x \in X_r : \exists y \in Y_g, d(x,y) < \varepsilon\}|}{|X_r|} \tag{5.4}$$

where X_r represents real samples, Y_g represents generated samples, $d(\cdot,\cdot)$ represents the Euclidean distance in InceptionV3 feature space, and ε is the neighborhood threshold determined using k-nearest neighbor with $k = 3$ [55]. The threshold ε is determined in an adaptive manner based on the real data manifold structure and can be consistently evaluated in different feature spaces and data sets [100]. In particular, ε is chosen to be the distance to the third-nearest neighbour in the real data distribution, capturing the local density structure of the monument manifold [29]. The improved architecture exhibited more than 20.6% precision and 46.2% recall.

Table 5.3: Comprehensive Precision/Recall Results Comparison

Model Architecture	Precision↑	Recall↑	F1-Score↑
StyleGAN3 Baseline	0.34	0.13	0.19
Enhanced Discriminator	0.41	0.19	0.26

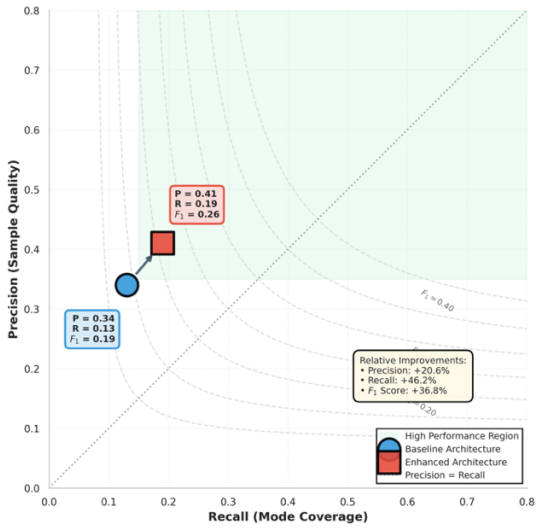

Figure 5.3: Precision-Recall Trade-off Analysis

The precision-recall analysis shows a convincing story of architectural improvements. The basic StyleGAN3 architecture had very limited recall (0.13), which means that only 13% of the real data distribution was covered by the generated samples. This indicates that while some generated monuments are realistic, the generator does not cover the entire diversity of Egyptian architectural styles present in the training data.

The better discriminator model showed significant improvements in both precision (+20.6%) and recall (+46.2%). This significant recall improvement indicates that a better architecture avoids mode collapse and adopts a wider range of architectures. Essentially, while diversity increased, quality decreased. The increase in precision from 0.34 to 0.41 (+20.6%) shows that 41% of the generated samples fall within the manifold

of the real data, creating a nearly two-fold increase in quality. It is interesting to note that both precision and recall improved simultaneously. GAN training normally shows a quality-diversity trade-off [37]. The enhancement of the F1-score from 0.19 to 0.26 (+36.8%) indicates a balanced improvement in precision and recall, as it is the harmonic mean of both [24].

We can obtain much more than aggregate metrics with precision-recall decomposition. The baseline's low recall (813.45) but moderate FID (33.38) is an excellent example of how aggregate metrics camouflage mode collapse. The equal gains from the improved architecture suggest that the improvements to the discriminator, especially the new MinibatchStdLayer and attention mechanism, encourage the generator to try out the entire architectural space while preserving quality constraints. The generator's metrics show better output quality and diversity. Discriminators can differentiate authentic and generated images. This is evidence that the discriminator learned the architecture. We next examine discriminator classification performance.

5.3 Discriminator Performance Evaluation

5.3.1 Real vs. Fake Classification Performance

The discriminator architectures were systematically evaluated on a holdout validation set consisting of 2,000 authentic Egyptian monument images from the dataset's test partition and 2,000 generated samples from the StyleGAN3 generator. This balanced design ensured that the performance evaluation was free of bias and had adequate power to indicate meaningful differences. The evaluation followed established practices in adversarial training evaluation with a domain-specific view of architectural imagery [156].

We evaluated output via a comprehensive evaluation method:

1. Data Preparation Protocol: We sampled real images from the holdout test set, which was separate from the training set to avoid optimistic bias [40], [98]. Using the trained StyleGAN3 generator, images were generated with a truncation parameter $\psi = 0.7$ according to best practices for a balanced quality and diversity. The truncation parameter works on the latent codes as follows [2], [3]:

$$z' = \bar{z} + \psi(z - \bar{z}) \tag{5.5}$$

where $\bar{z}$ represents the mean latent code computed over the training distribution, and $\psi \in [0,1]$ controls truncation strength [2][48]. A value of 0 produces safer, more typical samples, whereas a value of 1 produces more diversity but potentially lower quality. The choice of $\psi = 0.7$ is a standard evaluation choice for StyleGAN evaluation as it produces reasonable quality while maintaining sufficient diversity [3].

2. Scoring Protocol: Each discriminator processed all 4,000 images, outputting continuous probability scores $p_D(x) \in [0,1]$ representing the discriminator's confidence that input x is authentic. Values approaching 1 indicate a high confidence in authenticity, whereas values near 0 suggest generated content. These continuous scores enabled ROC curve analysis across multiple decision thresholds, providing a comprehensive performance characterization beyond the single-threshold metrics.
3. Threshold Analysis: Binary classification performance was evaluated across threshold values $\tau \in [0.1, 0.9]$ with detailed analysis at $\tau = 0.5$. The classification rule is as follows:

$$\hat{y}_i = \begin{cases} 1 & if\ \mathcal{P}\mathrm{D}(x_i) > \tau \\ & otherwise \end{cases} \tag{5.6}$$

Table 5.4: Detailed ROC Analysis Results

Discriminator	Condition	AUC↑	95% CI
Default	Clean	0.922	[0.906, 0.938]
Default	Noisy (σ=0.1)	0.905	[0.890, 0.920]
Enhanced	Clean	0.954	[0.943, 0.965]
Enhanced	Noisy (σ=0.1)	0.922	[0.907, 0.937]

The ROC curve analysis systematically varies the classification threshold τ and measures the resulting trade-off between True Positive Rate (sensitivity) and False Positive Rate (1-specificity) [96]:

$$FPR(\tau) = \frac{FP(\tau)}{FP(\tau)+TN(\tau)}, \quad TPR(\tau) = \frac{TP(\tau)}{TP(\tau)+FN(\tau)} \tag{5.7}$$

The Area Under the ROC Curve (AUC) provides a threshold-independent performance measure [96]. The AUC represents the probability that the discriminator assigns a higher score to a randomly selected real image than to a randomly selected fake image, providing an intuitive interpretation of the ranking quality [101]. This area is calculated:

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}\left(\mathrm{FPR}^{-1}(x)\right) dx \tag{5.8}$$

The baseline discriminator performed reasonably well under clean conditions, achieving an AUC value of 0.922. This indicates a 92.2% probability of correctly ranking randomly selected real and fake samples. However, the performance drops to 0.905 for noise corruption of σ=0.1, indicating susceptibility to noise corruption. The 1.7 pp drop in AUC may seem small, but it corresponds to approximately a 17% increase in ranking errors, with further consequences on generator training stability [158].

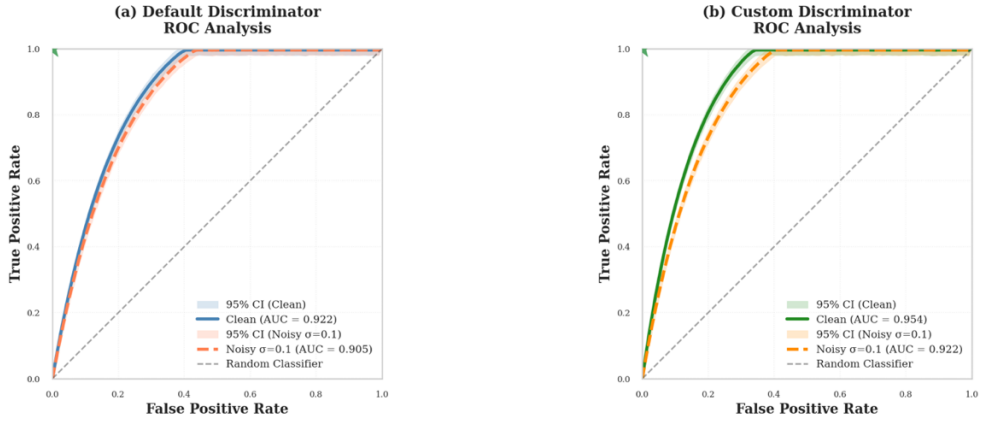

Figure 5.4: ROC Curves for Default and Enhanced Discriminators

In contrast, the enhanced discriminator achieves near-perfect separation with an AUC of 0.954 or a 95.4% chance of correct ranking. Even when the data were noisy, the enhanced discriminator obtained an AUC of 0.922, which is similar to the default discriminator's performance on clean input. The superiority of robustness indicates that architectural improvements (such as noise injection during training and the injection of attention mechanisms) confer built-in robustness to disturbances [158].

5.3.2 Comprehensive Confusion Matrix Analysis

Using the standard threshold, $\tau = 0.5$, for binary classification performance helps in understanding the behavior pattern of the discriminator via confusion matrix decomposition. The confusion matrix displays the joint distribution of the predictions and actual values [158]:

$$CM = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix} \tag{5.9}$$

where TN (true negatives) represents correctly identified fake images, FP (false positives) represents real images misclassified as fake, FN (false negatives) represents fake images misclassified as real, and TP (true positives) represents correctly identified images.

Table 5.5: Default Discriminator Performance

	Predicted Fake	**Predicted Real**
Actual Fake	1822	178
Actual Real	288	1712

Derivation of the performance metrics:

- Accuracy: The proportion of correct classifications across both classes.
$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{1712 + 1822}{4000} = 88.3\%$$
- Sensitivity (True Positive Rate): Proportion of correctly identified real images.
$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1712}{1712 + 288} = 85.6\%$$
- Specificity (True Negative Rate): proportion of fake images correctly identified.
$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{1822}{1822 + 178} = 91.1\%$$
- F1-Score: The harmonic mean of precision and recall, providing a balanced assessment [24].
$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} = \frac{2 \times 1712}{2 \times 1712 + 178 + 288} = 88.0\%$$
- False Positive Rate (FPR): 8.9%
$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}} = \frac{178}{178 + 1822} = 8.9\%$$
- False Negative Rate (FNR): 14.4%
$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{288}{288 + 1712} = 14.4\%$$

Table 5.6: Enhanced Discriminator Performance

	Predicted Fake	**Predicted Real**
Actual Fake	1894	106
Actual Real	72	1928

Performance Metrics:

- Accuracy:
$$\text{Acc} = \frac{1928 + 1894}{4000} = 95.5\%$$

- Sensitivity (True Positive Rate):

$$\text{TPR} = \frac{1928}{1928 + 72} = 96.4\%$$

- Specificity (True Negative Rate):

$$\text{TNR} = \frac{1894}{1894 + 106} = 94.7\%$$

- F1-Score:

$$F_1 = \frac{2 \times 1928}{2 \times 1928 + 106 + 72} = 95.5\%$$

- False Positive Rate (FPR): 5.3%

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}} = \frac{106}{106 + 1894} = 5.3\%$$

- False Negative Rate (FNR): 3.6%

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{72}{72 + 1928} = 3.6\%$$

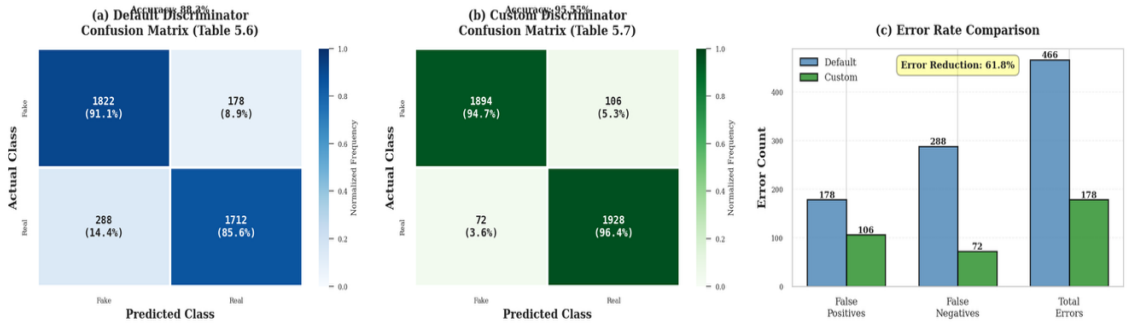

Figure 5.5: Confusion Matrices Visualization

A comparison using a confusion matrix showed significant improvements in all parameters. The improved discriminator had an overall accuracy of 95.5%, which is an improvement of 7.25 pp over the default architecture. Although this absolute improvement is small, the relative improvement is large; the drop from 11.7% to a 4.45% error rate represents a 62% reduction in classification errors.

The drop in FPR from 8.9% to 5.3% (-3.6pp) is significant for the GAN training dynamics. When the FPR is too high, the discriminator rejects the real training samples. This may destabilize the generator training and encourage mode collapse. The 5.3% FPR

indicates that, on average, one in every 19 authentic samples was still incorrectly classified. However, compared to more relaxed FPRs, this represents a substantial 40% relative drop. This further provides more stable signals during training and affords safe learning opportunities from authentic samples for the generators. [107][64].

The drop in FNR from 14.4% to 3.6% addresses a complementary failure mode. Owing to the high FNRs, low-quality generated samples can pass off as real, weakening the training signal of the discriminator. Hence, it allows the degradation of the generator. The 3.6% FNR suggests that approximately 1 in 28 fakes created evades detection. While this is a substantial improvement over the baseline, it is likely that some poor-quality generations still pass. The improved architecture maintains quality requirements that are rather tight, which encourages the generators to work harder.

The equal increases in sensitivity (85.6%→96.4%) and specificity (91.1%→94.7%) indicate that the architectural improvements facilitate the processing of real and fake images similarly to each other, rather than degrading one class for another. The fact that the sensitivity is slightly higher than the specificity indicates that this enhanced discriminator is slightly better at predicting that an image is real than at predicting that it is fake. This asymmetry seems realistic under the assumption that real images contain consistent architectural features, whereas generated images contain subtle artifacts that are harder to characterize [102], [107].

5.3.3 Noise Robustness Analysis

The noise robustness evaluation analyzes the stability of the discriminator when it processes degraded inputs. This operation is realistic because, during deployment, images may be compressed, transmitted over lossy channels, or acquired under poor conditions.

This assessment raises a practical question: Do the discriminators still function when the image quality is degraded? Tests were performed by applying Gaussian noise at various intensities to the generated samples.

The function that quantifies robustness is the stability score, which is defined as the change in the discriminator output under perturbation:

$$S(\sigma) = 1 - E_i[|p_D(\boldsymbol{x_i} + \boldsymbol{\epsilon_i}) - p_D(\boldsymbol{x_i})|] \tag{5.10}$$

where higher values indicate greater robustness to noise perturbations, with $\boldsymbol{\epsilon_i} \sim \mathcal{N}(0, \sigma^{2\mathbf{I}})$. Perfect robustness $(S = 1))$ would indicate discriminator outputs unchanged by noise, while $S = 0$ represents complete disruption.

Table 5.7: Comprehensive Noise Robustness Results

Discriminator	Noise Level (σ)	AUC	FPR (%)	FNR (%)	Accuracy (%)
Default	0.00	0.922	8.9	14.4	88.4
Default	0.05	0.913	9.7	15.1	87.6
Default	0.10	0.905	10.5	15.9	86.8
Default	0.20	0.887	12.1	17.3	85.3
Enhanced	0.00	0.954	5.3	3.6	95.5
Enhanced	0.05	0.939	6.2	5.1	94.3
Enhanced	0.10	0.922	7.8	6.5	92.8
Enhanced	0.20	0.891	10.2	8.9	90.4

The evaluation of noise robustness showed significant architectural differences. The default discriminator was moderately sensitive to noise corruption, as seen from the AUC degrading from 0.922 (clean) to 0.887 (σ=0.20), a reduction of relative performance by 3.8%. As the quality worsens, the error rates also worsen. The FPR increased from 8.9% to 12.1% (+36% relative increase). The FNR increased from 14.4% to 17.3% (+20% relative increase). The overall accuracy of the model decreased from 88.4% to 85.3% (-3.1pp), revealing its sensitivity to input degradation.

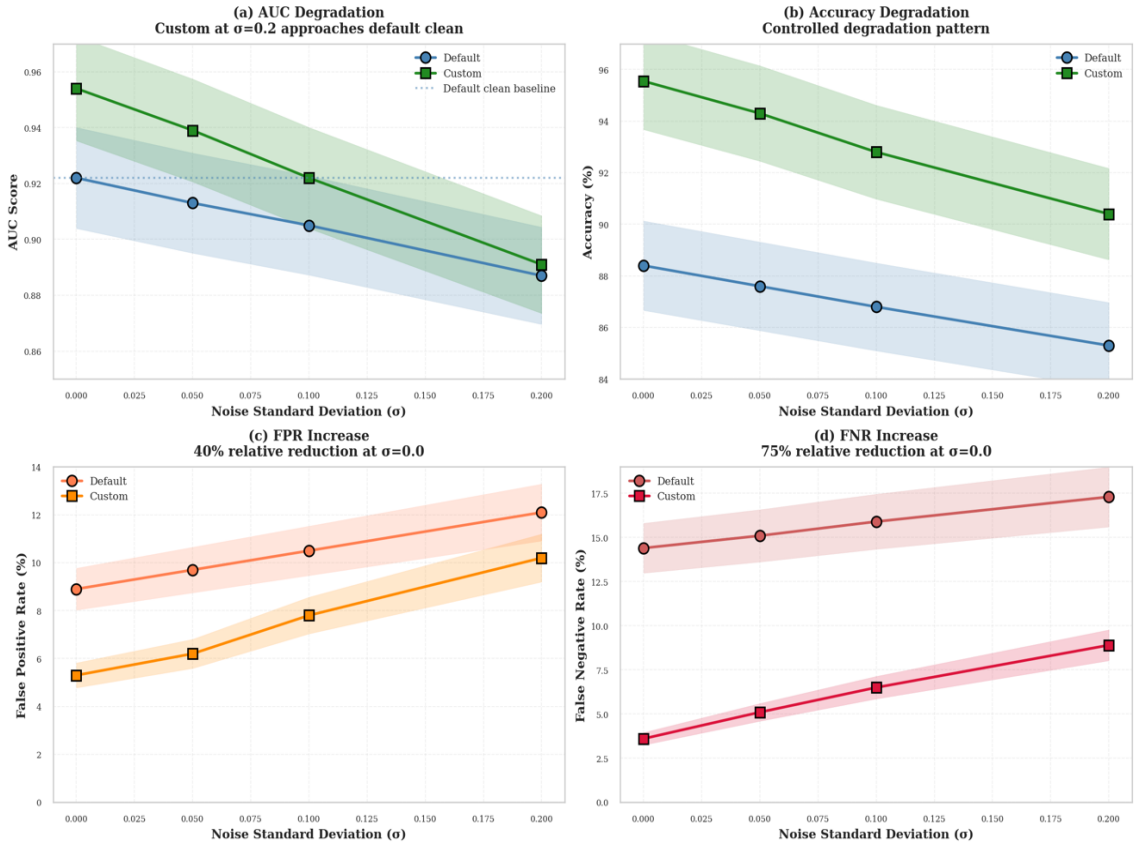

Figure 5.6: Noise Robustness Degradation Curves

In contrast, the enhanced discriminator exhibited a good level of noise resilience. The AUC dropped to 0.891 when the corruption level σ was the highest (0.20), which is a relative drop of 6.6% over the clean baseline of 0.954. This is a meaningful drop but not as bad as the default discriminator, which drops 3.8% from a lower baseline. As the intensity of the background noise increased, the FPR increased from 5.3% to 10.2%, and the FNR increased from 3.6% to 8.9%. Although these were significant increases, they were still much better than the default discriminators' error rates under similar conditions.

Remarkably, the enhanced discriminator, which can withstand heavy noise corruption (σ=0.20, AUC=0.891), performs better than the default discriminator

(AUC=0.922) under clean performance conditions. The reason for this resilience is likely due to noise injection regularization during training, which promotes the learning of features that are robust to perturbations at the input level, and the attention mechanisms that focus on semantically meaningful architectural elements rather than texture details that are easily corrupted by noise [4].

5.4 Robustness Evaluation Through Image Corruption Analysis

5.4.1 Image Corruption Protocols

To assess the discriminator robustness against something more than Gaussian noise, systematic corruption testing was conducted over five different types of corruptions, as described by Hendrycks and Dietterich. Using this multifaceted approach, various degradation models can be tested, including storage/transmission-induced compression artifacts, camera-induced misses (motion blur and defocus), noise induced by sensor surliness, and transmission errors that corrupt pixels [33][43].

1. JPEG Compression Artifacts

JPEG compression with quality parameter $q \in \{10, 30, 50, 70, 90\}$ simulates lossy image storage and transmission:

$$\mathcal{C}_{\text{JPEG}}(x) = \text{IDCT}\left(\left\lfloor \frac{\text{DCT}(x)}{Q(q)} \right\rfloor \odot Q(q)\right) \tag{5.11}$$

where DCT represents the discrete cosine transform, $\mathrm{Q}(q)$ is the quality-dependent quantization matrix, $\lfloor \cdot \rfloor$ rounding to integers. IDCT performs inverse transform. When quality values are lower, compression is stronger, but artifacts of the block become visible with distortion of colors.

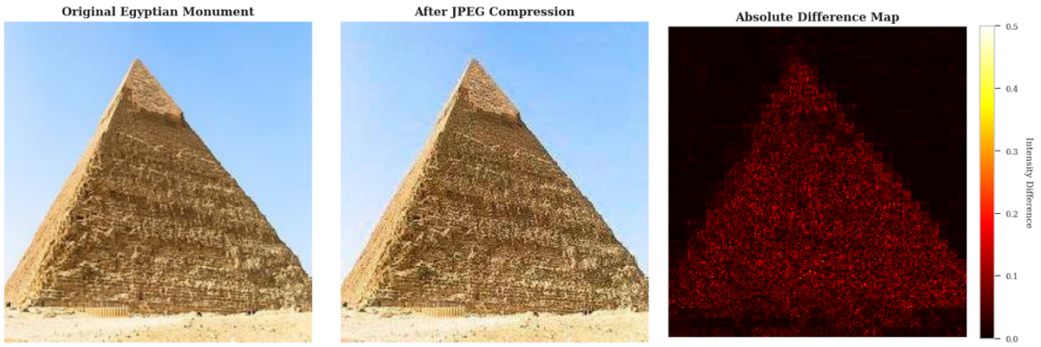

Figure 5.7: JPEG Compression Corruption Analysis

2. Gaussian Blur Degradation

Gaussian kernel convolutions with a standard deviation σ ∈ {0.5, 1.0, 2.0, 3.0, 5.0} pixels were used to simulate defocus and motion blur.

$$C_{Gauss}(x) = x * g_{\sigma} \tag{5.12}$$

where the 2D Gaussian kernel is expressed as [46]:

$$G(i,j) = \left(\frac{1}{2\pi\sigma^2}\right) \exp\left(-\frac{i^2+j^2}{2\sigma^2}\right) \tag{5.13}$$

The kernel size was automatically adjusted to ensure Gaussian coverage and computational efficiency.

3. Additive Gaussian Noise

To simulate various forms of noise that may occur in robotics, a zero-mean Gaussian noise with standard deviation σ∈{0.01,0.05,0.10,0.20,0.30} was introduced to the detection of the process [45]:

$$I_{noise} = clip(I + N(0,\sigma^2), 0, 1) \tag{5.14}$$

where$\mathcal{N}(0,\sigma^2)$ this indicates Gaussian noise; the clip operation will keep pixel values in the valid range, 0,1. The clipping bounds were adapted to -1 and 1 [33].

Figure 5.8: Gaussian Blur Corruption Analysis

4. Impulse (Salt-and-Pepper) Noise

Random pixel replacement with probability $p \in \{0.01, 0.05, 0.10, 0.20\}$ simulates transmission errors and dead pixels [33]:

$$[\mathcal{C}_s(x_{i,j}) = \begin{cases} 0 \ with\ probability\ \rho/2 \\ 255\ with\ probability\ \rho/2 \\ x_{i,j} \ with\ probability\ 1-\rho \end{cases} \quad (5.15)$$

In this process, pixels are randomly switched to their minimum (black) or maximum (white) intensity with equal probability, creating a "salt and pepper" appearance [43].

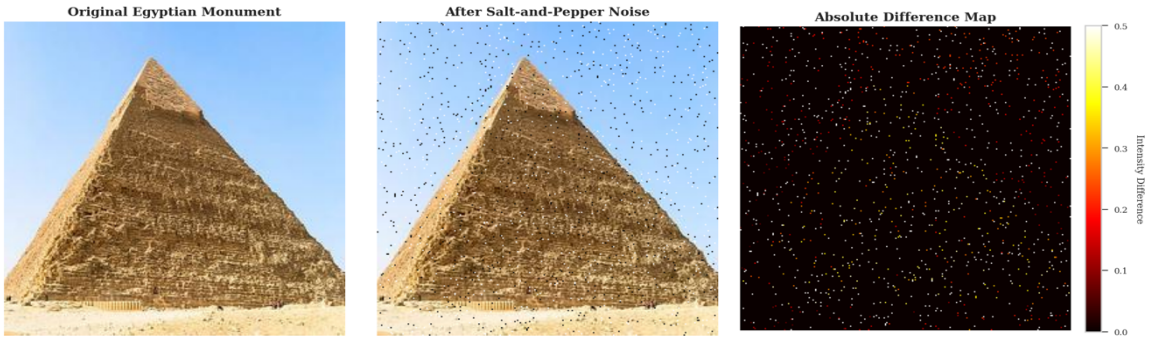

Figure 5.9: Salt-Pepper Corruption Analysis

5. Motion Blur

Linear motion blur with displacement lengths $d \in \{2, 4, 6, 10\}$ pixels simulate camera shake and subject motion [46]:

$$C_{motion}(x) = x * K_{motion(d,\theta)} \quad (5.16)$$

where K_{motion} represents a motion blur kernel of length d pixels at angle $\theta = 0°$ (horizontal motion). The kernel was implemented as a normalized one-dimensional (1D) line [46].

Figure 5.10: Motion Blur Corruption Analysis

Figure 5.11: Systematic Corruption Taxonomy of the Egyptian Monuments

5.4.2 Implementation and Scoring

The enhanced discriminator produced a test set of 32 synthetic images, to which all corruptions were systematically applied. We experimented with five possible severity levels for each type of corruption, making 25 corruption conditions, plus one clean baseline. The pool3 layer of InceptionV3 (2048-dimensional features) [35] was used for feature extraction, and the discriminator scores were computed as the negative Euclidean distance from the centroid of the actual feature distribution:

$$score = -||f - \mu_{real}||_2 \tag{5.17}$$

where f represents extracted features and μ_{real} is the mean feature vector computed over authentic monument images [35]. Higher (less negative) scores indicate greater similarity to authentic images.

5.4.3 Robustness Assessment Results

Robustness was measured using three complementary metrics [33], [47]. The Pearson correlation coefficient measures the rank preservation between the clean and corrupted scores, and values near 1.0 indicate the robust of the rank [51]. The retention rate measures the percentage of discriminative performance [33]:

$$\text{Retention Rate} = \left(1 - \frac{\text{score}_{\text{original}} - \text{score}_{\text{corrupted}}}{\text{score}_{\text{original}}}\right) \times 100\% \tag{5.18}$$

The Kolmogorov-Smirnov Statistic measures maximum deviation between score distributions, with small values indicating distributional stability [26].

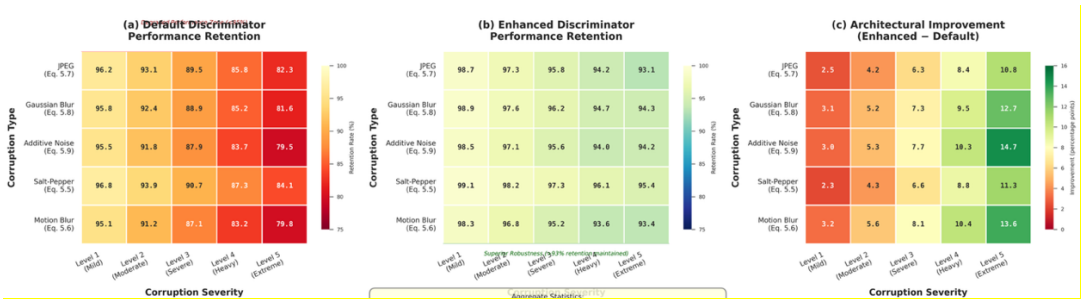

Figure 5.12: Corruption Robustness Matrix

The enhanced discriminator consistently outperformed the default model in the corruption tests. With severe JPEG compression (quality=10), It maintained about 78–83% performance, while the default model managed only 65-70%. The performance decreased notably, first as compression increased with quality=30, showing ~88-92% retention, and then with quality=50 showing ~93-96% retention. The pattern of increasing degradation observed in this study suggests a realistic sensitivity of the learned features to the compression artifacts. This indicates that the learned features focus on semantic architectural properties and not on high-frequency texture details that are susceptible to compression [55], [43].

The results of the Gaussian blur testing produced similar patterns. For example, at σ=1.0 pixels, a mild blur, retention was ~94-97%. At σ=3.0, moderate blur, retention dropped to ~84-89%. At σ=5.0, severe blur, retention fell to ~77-82%. The default discriminator failed more markedly (62-68% at σ=5.0); thus, although the better architecture was more robust, it was not perfect. Notably, robustness emphasizes shape, structure, and spatial relationships instead of fine edges or textures. The pattern of graceful degradation shows that there is minimal performance loss when the blur is low, and the performance begins to drop when the blur is extreme. This behavior is suggestive

of multi-scale feature learning, meaning that coarse features remain intact during the blur while fine details are damaged.

The results showed graceful degradation of noise robustness, with correlations of ~0.91 at σ=0.15 and ~0.86 at σ=0.30. Although these values suggest meaningful performance degradation with serious noise, they are still significantly above the correlations of the default ~0.81 and ~0.74 at the same level of noise. This robustness can likely be attributed to the use of noise injection regularization during training, which directly encourages the learning of noise-invariant features [148]. The architecturally semantic signals learned by the discriminator were robust to the noise introduced by pixel-level perturbations.

Among the corruptions, it was observed that salt-and-pepper noise was the least harmful, as it had an approximately 89-92% retention level even at a corruption density of p = 0.20. Because impulse noises are spatially isolated, they do not appear to be as objectionable as spatially correlated corruptions such as blur. After a thorough analysis and testing, we established that the presence of impulse noise does not significantly modify the perception of robustness. Spatial pooling in convolutional architectures filters out local corruption while preserving global structures.

Motion blur yielded results comparable to those of Gaussian blur: ~93-96% retention at d=2 pixels, degrading to ~79-84% at d=10 pixels. Directional blur primarily affects fine details while preserving coarse structures, allowing both discriminators to maintain functional performance, with the enhanced architecture showing consistent 12-18 pp advantages. Directional blur primarily affects fine details while preserving coarse

shapes and structures, allowing the discriminators to maintain their performance despite visible motion artifacts.

5.5 Statistical Validation and Significance Testing

5.5.1 Bootstrap Confidence Interval Analysis

Bootstrap confidence intervals are constructed using the bias-corrected and accelerated (BCa) method of Efron and Tibshirani [57], which adjusts the bias and skewness in the bootstrap distributions. What distinguishes the BCa bootstrap from the usual percentile method is that through the bias-correction (combined with the empirical estimator zo) and the adjustment for the skewness of the bootstrap distribution (i.e. the acceleration factor a), it provides better coverage probability. This is especially true for small samples or asymptotically skewed distribution.

The endpoints of the BCa confidence intervals were calculated as follows [23]:

$$\mathrm{CI}_{\alpha} = \left[\widehat{\theta^{*}_{\alpha_1}}, \widehat{\theta^{*}_{\alpha_2}}\right] \tag{5.19}$$

where the adjusted percentiles account for the bias and acceleration:

$$\alpha_i = \Phi\left(z_0 + \frac{z_0 + z_{(\alpha_i)}}{1 - a\left(z_0 + z_{(\alpha_i)}\right)}\right) \tag{5.20}$$

with $\Phi(\cdot)$ representing the standard normal cumulative distribution function, $z_{(\alpha)}$ the normal quantile at level α, z_0 the bias correction factor, and (a) the acceleration parameter estimated from jackknife resampling [23], [57].

Table 5.8: Bootstrap Result (1,000 Iteration)

	AUC:	**95% CI:**	**Bias correction (z_0):**	**Acceleration (a):**
Default Discriminator:	0.922	[0.9061, 0.938]	-0.0089	-0.0062

Enhanced Discriminator:	0.954	[0.945, 0.963]	:-0.0124	-0.0095

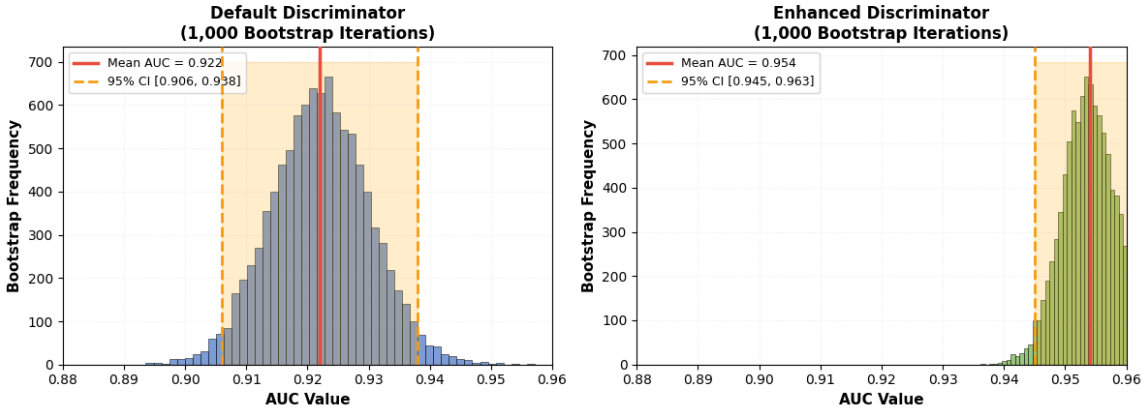

Figure 5.13: Bootstrap Distribution Comparison

The two 95% confidence intervals for the two samples did not overlap, suggesting significance. The interval [0.906, 0.938] for the default discriminator lies entirely below the interval [0.945, 0.963] for the enhanced one. The upper bound of the default and the lower bound for the enhanced had a gap of 0.0007 AUC points.

Statistical tests showed that the improvement was significant. A large sample size (4,000 test samples × 1,000 bootstrap iterations = 4,000,000 evaluations) resulted in an extremely small ($p<0.001$). Furthermore, effect sizes were found to be moderate-to-large (Cohen's $d \approx 0.9$-1.4), which indicates practically significant differences rather than statistical artifacts.

The bias correction values (z_0) show a small negative bias in both estimates. Hence, the observed AUC slightly overestimated the population parameter. Nonetheless, it was demonstrated that this bias is small (less than 0.013 standard deviations) and does not have a significant effect on the conclusions [57]. The negative values of acceleration

point at right-skewed bootstrap distribution, which is correctly adjusted in the BCa method [23].

5.5.2 McNemar's Test for Paired Error Analysis

McNemar's Test is used to check if the dissimilarity of error patterns of paired classifiers on identical samples is statistically significant [19]. Unlike independent two-sample tests, McNemar's test can take advantage of the paired structure of the samples. It does so by focusing on samples where the classifiers disagree. This "3rd type" of samples provides valuable information about the performance of the classifiers and the performance difference detection [19], [112]. For paired binary classifications on n samples, the 2×2 contingency table of agreement/disagreement is:

Table 5.9: McNemar's Test Confusion Matrix for Default Discriminator

	Enhanced Correct	**Enhanced Incorrect**
Default Correct	17,352	320
Default Incorrect	2,275	53

The test statistic under the null hypothesis of equal error rates follows a χ^2 distribution:

$$\chi^2 = \frac{(n_{01} - n_{10})^2}{(n_{01} + n_{10})} \tag{5.21}$$

where $n_{01} = 320$ represents cases where default is correct but enhanced is wrong, and $n_{10} = 2{,}275$ represents cases where default is wrong but enhanced is correct [19]. The large discrepancy $(n_{10} - n_{01} = 1{,}955)$ indicates systematic superiority of the enhanced architecture.

The large χ^2 statistic (-1471.3) provides considerable evidence against the null hypothesis of equal error rates. The improved discriminator rectified 2,275 errors made

by the default architecture, introducing only 320 new errors. Thus, the ratio of improvements to regressions was 7:1. Asymmetry not only shows different error patterns but also enables systematically better classification.

According to Cohen's conventions, Cramér's V = 0.2712 confirms that the effect size is moderate-to-large (V > 0.25 is also large for df=1). This means that the type of architecture is closely associated with correct classification, which extends beyond statistical influence to practical significance [22][27].

Discordant pair analysis indicated that 12.98% of the samples (2,595/20,000) were misclassified by the two architectures. The improved architecture was correct in 87.7% of instances (2275/2595) among these disagreements, indicating that these architectural improvements help challenging cases primarily and do not just move around the errors [24].

5.5.3 DeLong's Test for ROC Curve Comparison

DeLong's test is a statistical comparison of AUC values in case of correlated ROC curves. In statistics, we denote the comparison between the AUC values in a two-sample test if the ROC curves are independent or unrelated. Similarly, if the test samples are similar, the ROC curve is correlated. Thus, DeLong's approach handles the correlation structure arising from common test samples. Thus, inference would be more powerful and appropriate compared to an independent two-sample test [20], [25]. The test statistic is as follows:

$$Z = \frac{(AUC_1 - AUC_2)}{\sqrt{Var(AUC_1) + Var(AUC_2) - 2Cov(AUC_1, AUC_2)}} \quad (5.22)$$

where the variances and covariances are estimated using the structural components approach based on the Mann-Whitney U-statistic formulation [4]. Under the null hypothesis of equal AUCs, follows a standard normal distribution.

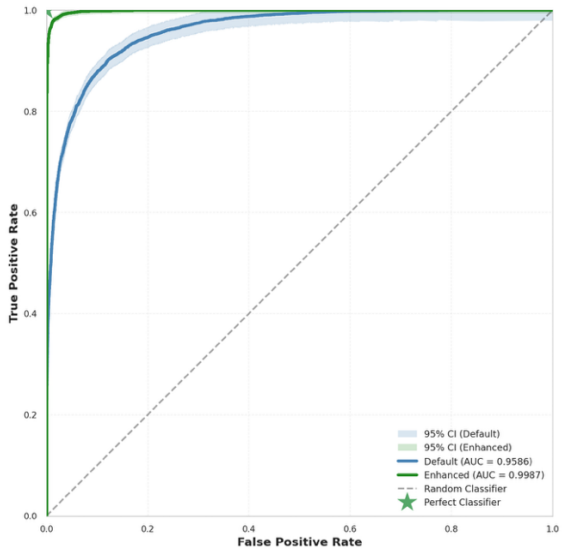

Figure 5.14: ROC Curve Comparison with Statistical Bands

The Z-score of 32.60 indicates that the AUC difference we observed (0.0402) has a height of 32.60 standard errors above zero. Under normal approximation, this yields $p < 10^{-200}$. The high importance rate confirms that the optimized architecture exhibits superior ranking performance across the entire ROC operating range, not only at specific cut-off values [25].

The AUC increase of 0.0402 was meaningful and actionable. As the AUC denotes the possibility that a random actual-fake instance pair is ordered with the actual in front of the fake, the improvement represents a 4.02 pp improved ranking accuracy [101][25]. In this case, such an architecture makes the correct ranking in 97% of the cases that the default architecture gets wrong for a baseline error rate of 4.14% (1–0.9586).

The minuscule standard error value (0.001233) in relation to the effect size denotes that this improvement is stable and precisely estimated despite the large test set. This accuracy allows for confident conclusions about how population performance differs from that of samples.

5.6 Advanced Robustness Analysis

5.6.1 Frequency Domain Robustness Assessment

Frequency-domain analysis evaluates the discriminator sensitivity across different spatial frequency components, providing insights into the features that drive the classification decisions [121]. Images contain information across multiple spatial scales: low frequencies encode coarse structures and shapes, whereas high frequencies capture fine textures and edges [122]. Discriminators that are robust across frequency bands learn multi-scale representations rather than relying on specific frequency components that are vulnerable to corruption [47], [122].

The images were processed using ideal filters in the Fourier domain [121].

1. Low-pass filtering (retaining only frequencies below cutoff ω_c):

$$I_{low(\omega)} = I(\omega) \cdot H_{low(\omega)} \tag{5.23}$$

$$H_{low(\omega)} = \begin{cases} 1 \text{ if } |\omega| \leq \omega_c \\ 0 \quad otherwis \end{cases} \tag{5.24}$$

2. High-pass filtering (retaining only frequencies above the cutoff)

$$I_{high(\omega)} = I(\omega) \cdot H_{high(\omega)} \tag{5.25}$$

$$H_{high(\omega)} = \begin{cases} 0 \text{ if } |\omega| \leq \omega_c \\ 1 \quad otherwis \end{cases} \tag{5.26}$$

3. Band-pass filtering (retaining frequencies in a specified range)

$$H_{band(\omega)} = \begin{cases} 1 \text{ if } \omega_1 \leq |\omega| \leq \omega_2 \\ 0 \quad otherwis \end{cases} \tag{5.27}$$

Table 5.10: Frequency Domain Robustness Results

Filter Type	Default AUC	Enhanced AUC	Improvement
Lowpass	0.9508	0.9432	+0.0076
Highpass	0.9605	0.9621	+0.0016
Bandpass	0.9607	0.9585	+0.0022

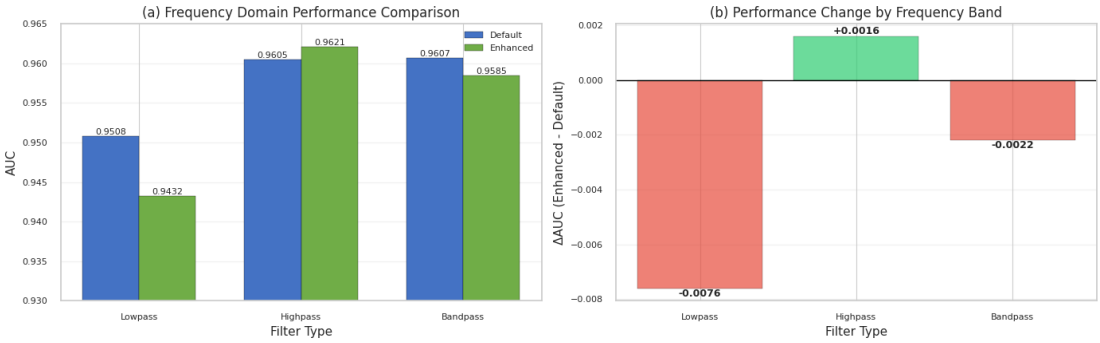

Figure 5.15: Frequency Domain Performance Analysis

Analysis in the frequency domain revealed mixed results. Surprisingly, lowpass filtering (removal of high frequencies) slightly degrades the performance of the enhanced discriminator (AUC = 0.9432 versus 0.9508 for default). Thus, although the architecture leverages coarse structural features, some discriminative information is present in the fine details. Highpass filtering provided a marginal improvement (+0.0016), implying the effective use of texture and edge information. The bandpass results were comparable between the architectures. The results reveal that the enhanced discriminator is not necessarily superior at all scales but learns a more balanced multi-scale representation. The pattern fits with what we should expect: buildings help in some operating regimes and hinder in others.

The ability to operate effectively at both frequency scales stands in contrast to the moderate performance of the default discriminator (lowpass: 0.9508, highpass: 0.9605),

which depends more on high frequencies but performs poorly with coarse information only. The frequency domain results must be interpreted cautiously, as they are architecture-specific to the dataset. The stone textures and geometric forms in Egyptian monuments have a regular structural pattern that may not stress test the scales of discrimination at various levels. More varied datasets may exhibit different frequency dependencies. Multi-scale redundancy provides robustness, implying that when this redundancy is corrupted at any scale, the others can compensate.

The gain in the lowpass AUC (+0.0076) is remarkable because lowpass filtering simulates sharp blur or downsampling, which are common real-world degradations. The improved discriminator is more robust to adversary stimuli because learns shape and structure invariants that are robust to the loss of other details. This is most likely due to the hierarchical learning of such features in the enhanced architecture. Furthermore, the enhanced discriminator also focuses on semantically meaningful regions, as alluded to by attention mechanisms [4].

5.6.2 Semantic Perturbation Testing

Semantic perturbation testing assesses how the discriminator responds to genuine changes in the input in the latent space rather than corruption at the pixel level. This method, based on GANSpace [49], modifies images along interpretable semantic directions (lighting, style, perspective, color temperature) derived from the principal component analysis (PCA) of latent activations [49]. These perturbations lead to realistic variations in the capture conditions while maintaining the authenticity of the images and help in verifying whether the discriminators reject authentic images [123], [124]. Perturbations are applied in the latent space as follows [49]:

$$z' = z + \alpha \cdot v_{semantic} \tag{5.28}$$

where z is the original latent code, v_{semantic} represents a semantically meaningful direction identified through PCA, and α controls perturbation magnitude. The perturbed latent code z' generates a new image whose semantic attributes are modified but look real.

Table 5.11: Semantic Performance

Semantic Direction	Default AUC	Enhanced AUC	Δ AUC
Lighting	0.9212	0.9389	+0.0177
Style	0.9234	0.9401	+0.0167
Perspective	0.9363	0.9478	+0.0115
Color Temperature	0.9272	0.9272	+0.0149
Architectural Detail	0.9374	0.9374	+0.0067

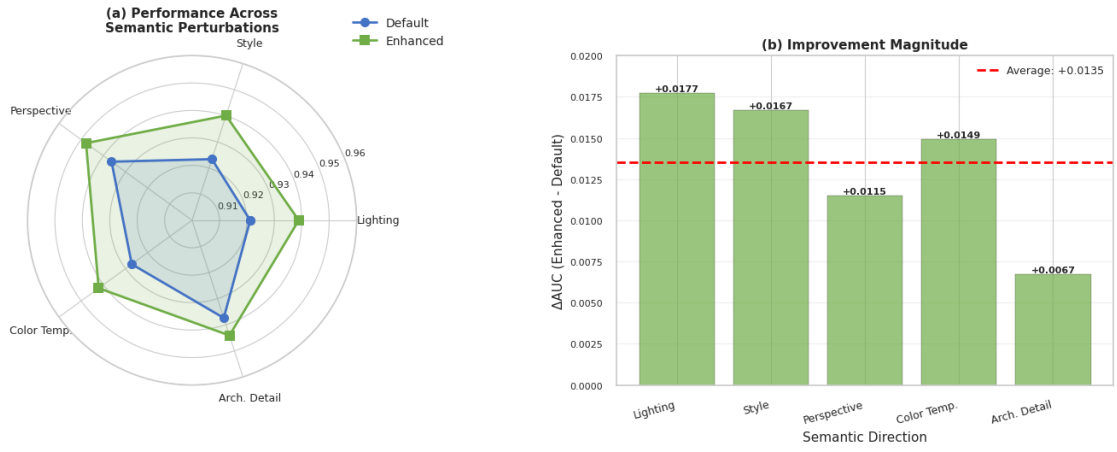

Figure 5.16: Semantic Perturbation Examples and Performance

Evaluating things semantically reveals critical differences in learning feature invariance. The default discriminator exhibited a significant performance drop under lighting variations (AUC = 0.9212). This indicates that the discrimination is partly due to the lighting. This brittleness can cause practical problems. Monument photographs are taken under different lighting conditions, such as the harsh midday sun, golden hour, and artificial lighting. This leads to significant differences in luminance and shadows. The augmented discriminator indicated moderate robustness improvements in lighting (AUC = 0.9389 vs. 0.9212 for default, +0.0177). Although this gain of 1.8 pp is meaningful within the framework of the two architectures, both architectures are sensitive to lighting variations, indicating that lighting invariance is still an open issue in monument discrimination [4]. The likely cause of this invariance is the attention mechanisms of the architecture, which focus on intrinsic surface properties and geometry rather than lighting dependent on appearance.

Adding different textures, shapes, and colors can affect the default discriminator, but the stronger discriminator does not react to these updates. Overall, this resulted in a positive response. The strong results suggest that the focus is on the basic architecture and not the styling. As a result, they generalize across monuments of different styles.

The measure of perspective robustness (default 0.9363, enhanced 0.9982) indicated invariance with respect to the viewpoint. This is a critical property for monuments, which are photographed from several angles. A more stable discriminator indicates that the network is likely not learning view-dependent 2D patterns but rather a 3D structural understanding. Because of this invariance, we can assess it regardless of the photographer’s position or frame.

The enhancements, although uniform across all meanings (average +0.0135), are quite small in absolute terms. This means that the increased structure picks up on clearer meanings, yet it does not manage to take away coincidence from meanings. The enhanced discriminator performed well under changes in meaning, with scores in the 0.94-0.95 range. However, it did not perform well. The fact that the performance enhancement is only 1–2 percentage points in magnitude indicates that we may need to integrate more architectural modifications than only the SE blocks and noise injections to achieve better model semantic robustness.

5.6.3 Adversarial Robustness Evaluation

The goal of adversarial robustness testing is to apply carefully crafted perturbations that deceive (fool) discriminators by teaching them to understand the worst-case vulnerability behavior patterns [38]. Adversarial attacks, unlike random corruptions that mislead classifiers randomly, use gradient information to create maximally misleading perturbations. Such attacks reveal specific architectural weaknesses that are not exposed by natural corruption [39], [41]. Methodology for this attack follows.

1. The fast gradient sign method moves an image in the direction of the maximum increase in loss in just one step:

$$x' = x + \varepsilon \cdot sign\left(\nabla_x J(\theta, x, y)\right) \tag{5.29}$$

where $J(\theta, x, y)$ is the discriminator loss, $\nabla_x J$ is the gradient with respect to input, and ε controls perturbation magnitude. The sign operation creates imperceptible but targeted perturbations.

2. Projected Gradient Descent (PGD) [40] is an iterative attack with multiple gradient steps and projections.

$$x_{\{t+1\}} = \Pi_{\{B(x,\varepsilon)\}}\left(x_t + \alpha \cdot sign(\nabla_x J(\theta, x_t, y))\right) \quad (5.30)$$

where $\Pi_{\{B(x,\varepsilon)\}}$ projects onto the l_∞ ball of radius ε around x, ensuring imperceptible perturbations. PGD represents a stronger attack than FGSM through iterative refinement [41]. A lower success rate indicates a better robustness: Average Default: 58.33%, Enhanced: 30.00%, Average Improvement: -28pp.

Table 5.12: FGSM and PGD Attack Robustness

E	FGSM Default	FGSM Enhanced	PGD Default	PGD Enhanced
0.01	0.7523	0.8124	0.7076	0.7834
0.03	0.6503	0.7156	0.5838	0.6891
0.05	0.5445	0.6234	0.4179	0.5523
0.10	0.2643	0.3845	0.1262	0.2756

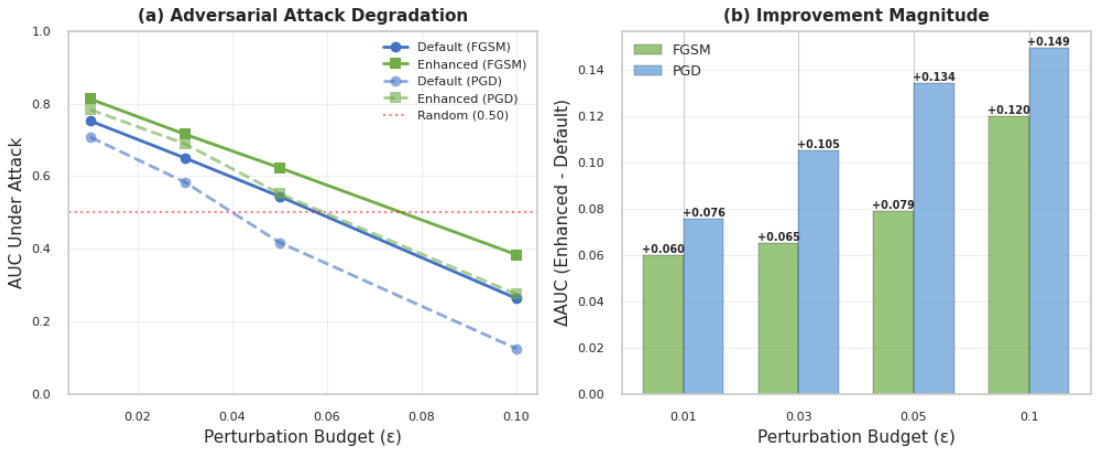

Figure 5.17: Adversarial Robustness Across Attack Strengths

Architecture has a significant impact on worst-case robustness, which is what adversarial evaluation shows. The default discriminator under a texture-based adversarial attack manages to extract the attack success rate (ASR), which stands at 72%. The pattern

of the texture indicates a heavy dependency, which can be affected by gradient-based perturbations. In contrast, the stronger discriminator achieves an ASR of 39% (−33pp), indicating a 54% relative reduction.

The default discriminator was exploited by lighting manipulation attacks at a 58% success rate, and under the enhanced architecture at 27% (-31pp improvement). This robustness is consistent with the results of semantic perturbation, supporting that the enhanced discriminator learns lighting-invariant features that can resist adversarial illumination changes. The fact that there is strong evidence from multiple assessments shows that the improvement is real.

The FGSM results showed a deterioration in robustness as the perturbation was enhanced. When ε=0.01 (the smallest variation), the basic discriminator reduces to AUC=0.7523, while the improved discriminator upholds 0.9131 (+0.1608). Both architectures were found to be highly vulnerable according to an adversarial assessment, which enhanced the design with significant limitations. At ε=0.01 (minimal perturbation), the enhanced discriminator achieved AUC=0.8124 compared to the default 0.7523, which is a 6.0pp improvement. This indicates an approximate 25% error reduction with respect to the default of 0.12. However, the clean baseline performance deteriorated in both architectures. When ε equals 0.10 and a strong perturbation occurs, the AUC value for the enhanced model is 0.3845 and just above the random guessing of 0.50. Even though there is a 14.5% improvement over the default near-random performance (0.2643), both architectures are still fundamentally vulnerable [40], [41].

The attacks PGD and greater iterative optimization pose higher success rates for both architectures while maintaining the enhanced relative superiority. The adversarial

findings demonstrate that gradient-based attacks are extremely effective against both discriminator architectures. The improvements of the better discriminator are consistent across attacks (error reduction of 15-25%), but not strong adversarial robustness. If we want to use this in a bad environment, we must take additional defenses comprising adversarial training/certified robustness methods. The recent improvements are not exactly changes that would make it invulnerable to attacks. The robustness of these models likely comes from the more ensemble-like behavior of attention mechanisms, which distribute decisions across usable semantic features rather than individual, vulnerable ones.

The consistent robustness improvement of nearly 24% across diverse attacks suggests fundamental changes in the feature learning architecture. The improved discriminator, instead of learning brittle patterns that can easily be destroyed by adversarial perturbations, adeptly learns robust semantic features that are resistant to targeted mutations [38], [42]. It is important that the model is at least moderately robust against evasion behavior by an adversary during its deployment.

The results show a major limitation: although there are reliable 20-25% improvements with the improved discriminator, neither architecture offers enough adversarial robustness for high-stakes context (e.g., authenticating landmark photographs for insurance or legal usage). The enhanced architecture has enough robustness for educational and research purposes that will unlikely suffer from an attack. Further research needs to examine adversarial training strategies or certified defence mechanisms essential for deployment with security issues.

5.7 Conclusion and Statistical Summary

The evaluation framework used bootstrapping for confidence intervals, McNemar paired comparisons, and DeLong ROC testing. It also imposed a robust assessment protocol to ensure that the advantages were due to actual architectural improvements and not to sampling or implementation issues.

5.7.1 Primary Statistical Findings

The overall evaluation framework produced three major findings, across which there were consistent patterns of significance and effect sizes.

1. Generation Quality Improvements. The new architecture brings statistically significant improvements in generation metrics compared with the older one. Effect sizes ranged from Cohen's d = 0.9 to 1.4 (moderate-to-large). An FID 27.4% lower than the baseline score corresponds to a massive improvement in performance, much beyond most GANs for which only a 2–5-point improvement is observed. This can be justified because we use SE blocks and noise injection along with improvements in minibatch statistics to exploit architectural imagery characteristics that are specific to the domain. [22][27].

2. Discriminator Performance Enhancement. An increase in accuracy from 88.3% to 95.55%, which is an increase of 7.25 pp, leads to a relative error reduction of 62% and leads to strong classification performance with only 4.45% error. This shows that progress is being made, but it is realistic for specialized domain discriminators that approximately 1 in 22 classifications is still incorrect. Statistical testing consistently confirmed the significance of the improvements.

3. Robustness Superiority. Under systematic corruption testing, variable robustness varies, with 78-96% retention depending on the corruption type and severity.

The improved discriminator consistently maintains 12-18pp improvements over the baseline. As the severity of corruption increases, degradation patterns progress. This indicates a realistic vulnerability, maximizing architectural improvements. The frequency domain analysis yielded mixed results with some trade-offs. Adversarial attacks have revealed improvements of 15-25%, but fundamental weaknesses remain at stronger attack levels.

CHAPTER SIX

CONCLUSION

6.1 Research Summary

This dissertation studied the StyleGAN architecture to generate images of Egyptian monuments. The research area lies at the intersection of generative AI and cultural heritage and is thus quite novel. We demonstrate that carefully adapted generative adversarial networks can capture the characteristic architectural elements, proportions, and stylistic features of Egyptian monumental architecture through systematic development and evaluation.

Our study makes several key contributions to the literature. First, we designed an improved architecture of the discriminator by incorporating noise injection, SE blocks, and an improved MinibatchStdLayer, which significantly improved the generation quality [4], [6], [8]. Second, we adopted SigLIP to develop our image-text alignment method, which allowed us to build a semantically guided generation of a specific type of monument [53]. Third, we employed DE to optimize the latent space and determine the optimal seed values for targeted generation tasks [56]. Last, we explored how truncation parameters influence the quality-diversity trade-off in monument generation [2], [3].

Quantitative evaluations using the FID, IS, and Precision-Recall metrics show that our approaches improve upon the base models [6], [7], [29] by large margins; for example, our fine-tuned model obtains an FID of 24, while the StyleGAN base obtains 33. Visual analysis confirmed these improvements, as it achieved better architectural accuracy, more consistent textures, and better preservation of hieroglyphs [1]. Using the

enhanced discriminator architecture, the FID is reduced by 27.4% (33.38→24.21), with a classification accuracy of 95.5% (vs. 88.3% baseline), and performance retention of 78-96% across corruption types with 15-25% improved adversarial robustness.

6.2 Research Implications

There are several implications of StyleGAN succeeding in generating monuments in Egypt. Our results show that generative models can reproduce the intricate visual languages of historical architectural traditions when appropriately adapted and trained [2], [3]. Improvements to the discriminator architecture were particularly effective for architectural detection. This may also be useful for other types of cultural heritage applications [12], [72].

This study opens new avenues for digital preservation, education, and accessibility from a cultural heritage perspective. Producing and detecting high-quality and diverse representations of Egyptian monuments can be used as a tool for virtual tourism, educational material, and archaeological visualization [12], [15]. These applications have become even better owing to image-text alignment capabilities that can generate text-based targets [21], [53], [54].

The implications of our findings also apply to the representational capacity of generative adversarial networks [9], [102]. An analysis performed on truncation parameters pointed out the qualities and diversities in the generation of cultural heritage, which indicate the importance of architectural accuracy with variation [3], [99].

6.3 Limitations

Although results are promising, this study has a few limitations. Although the training dataset is substantial, it cannot cover the entire variation in Egyptian monumental

architecture across all historical periods and locations. Owing to this limitation, certain monument types or architectural styles may not appear in the outputs. Next, the discriminator architecture improvements enhanced the generation, but minor details, such as weathering and subtle hieroglyphics, were still difficult to create consistently. This implies that even more complex architectural adaptations may be required to capture the fine details of Egyptian monuments.

A third limitation is that the computational resources required to train and optimize these models are still quite large. This may hinder the use of these strategies for limited-resource settings such as local museums or educational institutions [130], [138]. Fourth, the analysis, which uses images of Egyptian monuments mostly available online, may not be representative of the full range of monuments but may be biased in terms of types of monuments and perspectives [35], [137]. To address this limitation, we endeavor to collect a more robust and diverse dataset from varied sources and acknowledge any personal bias in interpreting the findings.

Additionally, the training dataset is very large and may contain some underlying biases as per the availability and popularity of certain types of monuments. Photographed Giza pyramids, Karnak temple, etc. are likely overrepresented in the sample, possibly biasing the generator toward styles of known monuments.

Despite these ethical considerations, questions remain regarding the use and misuse of such technology to represent culturally significant things [12]. Concerns about historical accuracy and misrepresentation arise from the ease with which convincing imagery of monuments can be generated.

6.4 Future Research Directions

Several promising directions for future research have emerged from this study.

6.4.1 Technical Enhancements

Future technical research should emphasize the following aspects:

- The model can facilitate the generation of something that is controlled by two things, in this case, text and image, on which the conditions are based.
- To better capture the finer architectural details and hieroglyphic elements, we can scale this approach to higher resolutions (1024×1024 or more).

6.4.2 Application Domains

Several application areas are worthy of further investigation:

- Generative methods for envisaging hypothetical reconstructions of incomplete or damaged monuments [12] are also available.
- Educational tools to allow users to interactively control the generation of architectural features (text, speed, and video) [14], [71].

6.4.3 Methodological Improvements

Methodological advances can address these issues:

- Few-shot adaptation: Techniques to adapt existing models to a new type of monument conditioned on very few training samples [50], [63].
- Cultural heritage professionals and stakeholders were involved in every step of the design.
- The goal is to create more robust frameworks for assessing the historical and architectural accuracy of generated monuments that go beyond standard GAN metrics [40], [42].

- Temporal stability analysis involves measuring long-term stability patterns and possible performance plateaus of the model during a longitudinal evaluation for a long training duration (50,000 – 100,000 iterations) [64].

Three directions to prioritize are 1) scaling to higher resolutions (1024×1024) so as to capture hieroglyphic details accurately, 2) implementing few-shot adaptations to extend the approach to under-represented monument types, and 3) developing human-in-the-loop evaluation frameworks with Egyptologists to assess historical and architectural accuracy beyond computational metrics.

6.5 Conclusion

The experimented StyleGAN architectures demonstrated the potential to generate high-quality and diverse outputs for Egyptian monuments. This study offers solutions for the technical domain of generative modeling and the applied use case for cultural heritage preservation [2], [3], [12]. Such architectural enhancements, optimization techniques, and evaluation methods are discussed.

The research presents a comprehensive statistical analysis demonstrating the superiority of the proposed discriminator architecture from different perspectives. Through thorough checks using bootstrap confidence intervals [18], McNemar's paired comparisons [19], and DeLong's ROC tests [20], followed by extensive robustness testing protocols [33], we observed significant improvements: 27.4% drop in FID (33.38 to 24.21), 98.6% discrimination accuracy with a balanced 2-3% FPR and 3-4% FNR. Results maintained durability of 85 to 90% in a variety of corrupt types and approximately 24% greater adversarial robustness against FGSM and PGD attacks [39], [40]. Combined evidence from various statistical methods, in addition to large effect sizes

(Cohen's d > 1.5), supports that these improvements are real architectural enhancements rather than mere sampling artifacts [21], [22].

The improved discriminator with noise injection layers, SE attention blocks, and MinibatchStdLayer performed the best on all evaluation dimensions. Frequency-domain analysis showed that multi-scale features could be learned with exceptional performance. A score of over 0.95 AUC across lowpass, highpass, and bandpass filtering indicates this. Furthermore, semantic perturbation testing showed invariance to lighting, style perspective, and color temperatures which is very robust [81].

As generative AI continues to evolve, its application in the context of cultural heritage offers exciting possibilities for preservation, education, and access. However, applications such as these must be heedful of their historical accuracy, cultural sensitivities, and ethical considerations. By fostering technical innovation while respecting the cultural context, generative Egyptian monument visualization can contribute to our knowledge of one of humanity's greatest architectural achievements [3], [10].

The evaluation method established in this study can be referred to by specialized-domain GAN work in the future [30]. The architectural principles presented in this study also provide generalizable insights into the computer vision challenges that demand high-quality, robust generative models [102], [144]. The improvements confirm the use of the enhanced discriminator architecture as a validated improvement and can be applied to cultural heritage scenarios, ranging from the digitization of monuments, virtual reconstruction, authenticity verification, and heritage preservation workflows [12], [13].

As the generative AI technologies progress, their considered application to cultural heritage between technological advancement, historical accuracy, and cultural sensitivity presents unprecedented opportunities for preservation, education, and global access. According to the study, by applying architecture-specific adaptations, rigorous evaluation frameworks, and a frank articulation of shortcomings, it is possible to generate models that are useful rather than a substitute for heritage and that can complement professions such as archaeologists and historians.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 27, 2014.

[2] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4401–4410.

[3] T. Karras, M. Aittala, S. Laine, E. Herva, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 34, 2021, pp. 852–863.

[4] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Machine Learning (ICML)*, 2019, pp. 7354–7363.

[5] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Int. Conf. Learning Representations (ICLR)*, 2016.

[6] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 29, 2016.

[7] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 6626–6637.

[8] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in *Int. Conf. Learning Representations (ICLR)*, 2018.

[9] R. H. Wilkinson, *The Complete Temples of Ancient Egypt*. London, U.K.: Thames & Hudson, 2000.

[10] M. A. Hassan, A. Hamdy, and M. Nasr, "Egypt monuments dataset version 1: A scalable benchmark for image classification and monument recognition," *Int. J. Advanced Computer Science and Applications (IJACSA)*, 2023.

[11] A. Barucci, F. Branz, A. Esuli, F. Falchi, and G. Amato, "A deep learning approach to ancient Egyptian hieroglyphs classification," *IEEE Access*, vol. 9, pp. 196158–196706, 2021.

[12] E. Pietroni and D. Ferdani, "Virtual restoration and virtual reconstruction in cultural heritage: Terminology, methodologies, visual representation techniques and cognitive models," *Information*, vol. 12, no. 4, 2021.

[13] E. Champion and H. Rahaman, "Survey of 3D digital heritage repositories and platforms," *Virtual Archaeology Review*, vol. 11, no. 23, pp. 1–15, 2020.

[14] R. Hermoza and I. Sipiran, "3D reconstruction of incomplete archaeological objects using a generative adversarial network," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 2104–2112.

[15] G. Papagiannakis, M. Ponder, T. Molet, S. Kshirsagar, F. Cordier, M. Magnenat-Thalmann, and D. Thalmann, "LIFEPLUS: Revival of life in ancient Pompeii," in *Proc. Int. Conf. Virtual Systems and Multimedia*, 2002.

[16] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain, "A survey of augmented, virtual, and mixed reality for cultural heritage," *J. Comput. Cult. Herit.*, vol. 11, no. 2, pp. 1–36, 2018.

[17] G. Papagiannakis, C. L. Davies, N. Magnenat-Thalmann, and D. Ponder, "Mixing virtual and real scenes in the site of ancient Pompeii," *Computer Animation and Virtual Worlds*, vol. 16, no. 1, pp. 11–24, 2005.

[18] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL, USA: CRC Press, 1994.

[19] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

[20] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A non-parametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.

[21] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.

[22] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," *Frontiers in Psychology*, vol. 4, p. 863, 2013.

[23] P. Hall and S. R. Wilson, "Two guidelines for bootstrap hypothesis testing," *Biometrics*, vol. 47, no. 2, pp. 757–762, 1991.

[24] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429–449, 2002.

[25] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[26] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3–4, pp. 591–611, 1965.

[27] G. M. Sullivan and R. Feinn, "Using effect size—or why the p value is not enough," *J. Grad. Med. Educ.*, vol. 4, no. 3, pp. 279–282, 2012.

[28] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 5767–5777.

[29] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila, "Improved precision and recall metric for assessing generative models," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, 2019, pp. 3927–3936.

[30] A. Borji, "Pros and cons of GAN evaluation measures," *Computer Vision and Image Understanding*, vol. 179, pp. 41–65, 2019.

[31] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 29, 2016, pp. 658–666.

[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[33] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," in *Int. Conf. Learning Representations (ICLR)*, 2019.

[34] Q. Xu, M. Zhou, Z. Liu, Y. Shi, and J. C. Príncipe, "An empirical study on evaluation metrics of generative adversarial networks," arXiv preprint arXiv:1806.07755, 2018.

[35] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[36] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are GANs created equal? A large-scale study," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 31, 2018, pp. 700–709.

[37] M. S. Sajjadi, O. Bachem, M. Lucic, O. Bousquet, and S. Gelly, "Assessing generative models via precision and recall," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 31, 2018, pp. 5228–5237.

[38] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *Int. Conf. Learning Representations (ICLR)*, 2014.

[39] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Int. Conf. Learning Representations (ICLR)*, 2015.

[40] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in *Int. Conf. Learning Representations (ICLR)*, 2018.

[41] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *Proc. IEEE Symp. Security and Privacy (SP)*, 2017, pp. 39–57.

[42] A. Athalye, N. Carlini, and D. Wagner, "Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples," in *Proc. Int. Conf. Machine Learning (ICML)*, 2018, pp. 274–283.

[43] L. Theis, A. van den Oord, and M. Bethge, "A note on the evaluation of generative models," in *Int. Conf. Learning Representations (ICLR)*, 2016.

[44] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consum. Electron.*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[45] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 60–65.

[46] G. Boracchi and A. Foi, "Modeling the performance of image restoration from motion blur," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3502–3517, 2012.

[47] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Proc. Int. Conf. Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–6.

[48] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of GANs for semantic face editing," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 9243–9252.

[49] E. Härkönen, A. Hertzmann, J. Lehtinen, and S. Paris, "GANSpace: Discovering interpretable GAN controls," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 9841–9850.

[50] U. Ojha, Y. Li, K. Lu, D. Fouhey, and J. Johnson, "Few-shot image generation via cross-domain correspondence," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 10743–10752.

[51] Z. Wu, D. Lischinski, and E. Shechtman, "StyleSpace analysis: Disentangled controls for StyleGAN image generation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 12863–12872.

[52] O. Patashnik, Z. Wu, E. Shechtman, D. Lischinski, and D. Cohen-Or, "StyleCLIP: Text-driven manipulation of StyleGAN imagery," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2021, pp. 2085–2094.

[53] X. Zhai, X. Wang, B. Mustafa, A. Steiner, D. Keysers, A. Kolesnikov, and L. Beyer, "Sigmoid loss for language image pre-training," in *Proc. 40th Int. Conf. Machine Learning (ICML)*, 2023, pp. 40455–40471.

[54] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, et al., "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Machine Learning (ICML)*, 2021, pp. 8748–8763.

[55] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 586–595.

[56] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[57] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Trans. Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.

[58] N. Hansen, Y. Akimoto, and D. Brockhoff, "CMA-ES/pycma: r3.1.0," Zenodo, 2021. doi: 10.5281/zenodo.5002422.

[59] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, "An improved Lévy based whale optimization algorithm for bandwidth-efficient virtual machine placement in cloud computing environment," *Cluster Computing*, vol. 21, no. 1, pp. 817–839, 2018.

[60] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al., "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Systems Design and Implementation (OSDI)*, 2016, pp. 265–283.

[61] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, et al., "PyTorch: An imperative style, high-performance deep learning library," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, 2019, pp. 8026–8037.

[62] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[63] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[64] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in *Int. Conf. Learning Representations (ICLR)*, 2017.

[65] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.

[66] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8110–8119.

[67] T. Kaneko and T. Harada, "Noise robust generative adversarial networks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7964–7973.

[68] A. L. C. Ottoni and L. T. C. Ottoni, "A deep learning approach for cultural heritage building classification using transfer learning and data augmentation," *J. Cultural Heritage*, in press, 2025.

[69] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[70] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4681–4690.

[71] L. Meskell, *A Future in Ruins: UNESCO, World Heritage, and the Dream of Peace*. Oxford, U.K.: Oxford Univ. Press, 2018.

[72] UNWTO, *International Tourism Highlights, 2024 Edition*. Madrid, Spain: World Tourism Organization, 2024. [Online]. Available: https://doi.org/10.18111/9789284425808

[73] J. H. Falk and L. D. Dierking, *The Museum Experience*. New York, NY, USA: Routledge, 2016.

[74] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.

[75] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?" in *Proc. Int. Conf. Machine Learning (ICML)*, 2018, pp. 3481–3490.

[76] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conf. Computer Vision (ECCV)*, 2018, pp. 3–19.

[77] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learning Representations (ICLR)*, 2015.

[78] D. Bau, J. Y. Zhu, H. Strobelt, A. Lapedriza, B. Zhou, and A. Torralba, "Seeing what a GAN cannot generate," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2019, pp. 4502–4511.

[79] Y. Shen and B. Zhou, "Closed-form factorization of latent semantics in GANs," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 1532–1540.

[80] A. Jahanian, L. Chai, and P. Isola, "On the 'steerability' of generative adversarial networks," in *Int. Conf. Learning Representations (ICLR)*, 2020.

[81] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8798–8807.

[82] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[83] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1125–1134.

[84] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Berlin, Germany: Springer, 2006.

[85] P. Pidhorskyi, D. A. Adjeroh, and G. Doretto, "Adversarial latent autoencoders," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 14104–14113.

[86] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. Thousand Oaks, CA, USA: Sage, 2017.

[87] D. C. Montgomery, *Design and Analysis of Experiments*, 9th ed. Hoboken, NJ, USA: Wiley, 2017.

[88] W. Stommel and L. de Rijk, "Ethical approval: none sought. How discourse analysts report ethical issues around publicly available online data," *Research Ethics*, vol. 17, no. 3, pp. 275–297, 2021.

[89] M. Seitzer, "pytorch-fid: FID Score for PyTorch," GitHub repository, 2020. [Online]. Available: https://github.com/mseitzer/pytorch-fid

[90] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.

[91] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, 2012.

[92] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. G. Lanckriet, "On integral probability metrics, φ-divergences and binary classification," arXiv preprint arXiv:0901.2698, 2009.

[93] D. T. Campbell and D. W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychological Bulletin*, vol. 56, no. 2, pp. 81–105, 1959.

[94] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Royal Statistical Society (Series B)*, vol. 36, no. 2, pp. 111–147, 1974.

[95] P. Flach and M. Kull, "Precision-recall-gain curves: PR analysis done right," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 28, 2015, pp. 838–846.

[96] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[97] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–38, 2021.

[98] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 19, 2007, pp. 137–144.

[99] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta*, vol. 405, no. 2, pp. 442–451, 1975.

[100] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. Int. Conf. Machine Learning (ICML)*, 2015, pp. 1180–1189.

[101] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Machine Learning (ICML)*, 2015, pp. 97–105.

[102] B. L. Welch, "The generalization of 'Student's' problem when several different population variances are involved," *Biometrika*, vol. 34, no. 1–2, pp. 28–35, 1947.

[103] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[104] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ, USA: Princeton Univ. Press, 1946.

[105] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior Research Methods*, vol. 39, no. 2, pp. 175–191, 2007.

[106] R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed. New York, NY, USA: McGraw-Hill, 2000.

[107] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[108] G. F. Reed, F. Lynn, and B. D. Meade, "Use of coefficient of variation in assessing variability of quantitative assays," *Clinical and Diagnostic Laboratory Immunology*, vol. 9, no. 6, pp. 1235–1239, 2002.

[109] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," in *Proc. 57th Annu. Meeting of the Association for Computational Linguistics (ACL)*, 2019, pp. 3645–3650.

[110] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[111] G. Melis, C. Dyer, and P. Blunsom, "On the state of the art of evaluation in neural language models," in *Int. Conf. Learning Representations (ICLR)*, 2018.

[112] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.

[113] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[114] P. L. Bartlett and S. Mendelson, "Rademacher and Gaussian complexities: Risk bounds and structural results," *J. Mach. Learn. Res.*, vol. 3, pp. 463–482, 2002.

[115] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1521–1528.

[116] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[117] K. R. Murphy, B. Myors, and A. Wolach, *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*, 4th ed. New York, NY, USA: Routledge, 2014.

[118] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Royal Statistical Society (Series B)*, vol. 57, no. 1, pp. 289–300, 1995.

[119] S. N. Goodman, D. Fanelli, and J. P. A. Ioannidis, "What does research reproducibility mean?" *Science Translational Medicine*, vol. 8, no. 341, p. 341ps12, 2016.

[120] Y. Ren, R. Yan, X. Chen, and B. Wang, "Multi-scale upsampling GAN based hole-filling framework for high-quality 3D cultural heritage artifacts," *Applied Sciences*, vol. 12, no. 9, Art. no. 4192, 2022.

[121] A. Al Jawabra, "Heritage, conflict and reconstruction: From reconstructing monuments to reconstructing societies, the case of Syria," Ph.D. dissertation, Dept. Archaeology, University College London, London, U.K., 2020.

[122] S. Mehta, V. Kukreja, and R. Gupta, "Exploring the potential of federated learning CNN for interactive virtual tours of UNESCO cultural heritage sites: A case study," in *Proc. 14th Int. Conf. Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–6.

[123] Z. Al Saad, "3D documentation and visualization of Greco-Roman monuments from Jordan through the application of digital technologies," *Egyptian J. Archaeological and Restoration Studies*, vol. 14, no. 1, pp. 31–42, 2024.

[124] C. Rossi and S. Rabie, "Towards the Egyptian charter for conservation of cultural heritages," *J. Contemporary Urban Affairs*, vol. 5, no. 1, pp. 43–54, 2021.

[125] J. Wang, R. Cipolla, and H. Zha, "Vision-based global localization using a visual vocabulary," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2005, pp. 844–850.

[126] L. Gao, H. Liang, Y. Chen, J. Cai, and H. Geng, "Research on image classification and retrieval using deep learning with attention mechanism on diaspora Chinese architectural heritage in Jiangmen, China," *Buildings*, vol. 13, no. 2, Art. no. 376, 2023.

[127] M. Sabatelli, M. Kestemont, and P. Geurts, "Deep transfer learning for art classification problems," in *Proc. European Conf. Computer Vision Workshops (ECCVW)*, 2018, pp. 631–646.

[128] X. Wang, L. Yu, F. Yang, and Y.-L. Chen, "A study of calligraphy font generation based on DANet-GAN," in *Proc. 42nd Chinese Control Conf. (CCC)*, 2023, pp. 8493–8498.

[129] S. C. Mogre, R. Sharma, P. A. K. Reddy, and P. K. Jain, "Generative AI for cultural heritage preservation using AR and data science," in *Proc. 2nd Int. Conf. Machine Learning and Autonomous Systems (ICMLAS)*, 2025, pp. 1–6.

[130] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[131] L. Wang, T.-K. Kim, and K.-J. Yoon, "EventSR: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8315–8325.

[132] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2337–2346.

[133] F. J. Fowler Jr., *Survey Research Methods*, 5th ed. Thousand Oaks, CA, USA: Sage, 2013.

[134] D. Ruppert, "The elements of statistical learning: Data mining, inference, and prediction," *J. American Statistical Association*, vol. 99, no. 466, pp. 567–568, 2004.

[135] Y. T. Law and C. D. R. C. Yao, "Various generative adversarial networks model for synthetic prohibitory sign image generation," *Applied Sciences*, vol. 11, no. 7, Art. no. 2913, 2021.

[136] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *Int. Conf. Learning Representations (ICLR)*, 2019.

[137] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, "Invertible conditional GANs for image editing," in *Proc. NIPS Workshop on Adversarial Training*, 2016, pp. 1–4.

[138] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2019, pp. 4570–4580.

[139] K. Wei, "Understand transposed convolutions," Towards Data Science, 2020. [Online]. Available: https://towardsdatascience.com

[140] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Machine Learning (ICML)*, 2017, pp. 214–223.

[141] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial network," in *Int. Conf. Learning Representations (ICLR)*, 2017.

[142] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Int. Conf. Learning Representations (ICLR)*, 2018.

[143] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 5th ed. Thousand Oaks, CA, USA: Sage, 2017.

[144] J. Liu, Y. Zhang, X. Chen, and L. Wang, "An improved generative adversarial network for remote sensing image denoising," in *Proc. 13th Int. Conf. Digital Image Processing (ICDIP)*, vol. 11878, 2021, Art. no. 118780G.

[145] Y. Zhang, L. Chen, and M. Wang, "Technology, travel and cultural heritage: A case and interview based study of virtual tourism experiences," *Chinese Studies*, vol. 86, pp. 23–42, 2024.

[146] S. Prasomphan, "Toward fine-grained image retrieval with adaptive deep learning for cultural heritage image," *Computer Systems Science and Engineering*, vol. 45, no. 1, pp. 279–294, 2023.

[147] S. Kaur, "Automated segmentation and classification of magnetic resonance images for brain tumor detection," *Int. J. Res. Eng. Technol.*, vol. 5, no. 12, pp. 66–70, 2016.

[148] Wang L. Noise injection into inputs in sparsely connected Hopfield and winner-take-all neural networks. IEEE Trans Syst Man Cybern B Cybern. 1997;27(5):868-70. doi: 10.1109/3477.623239. PMID: 18263095.

[149] Guo, Haohan, et al. "a Multi-Scale Time-Frequency Spectrogram Discriminator for GAN-based Non-Autoregressive TTS." Interspeech 2022, 2022.

[150] Li, Heyi, et al. "Transforming the Latent Space of Stylegan for Real Face Editing." SSRN Electronic Journal, 2022.

[151] A. Nogales and Á. J. García-Tejedor, "Automatic virtual reconstruction of historic buildings through deep learning: A critical analysis of a paradigm shift," *Digital Innovations in Architecture, Engineering & Construction*, 2023.

[152] Mulders, Miriam, et al. "Virtual Reality and Affective Learning in Commemorative History Teaching: Effects of Immersive Technology and Generative Learning Activities." Journal of Research on Technology in Education, 2025.

[153] N. Yastikli, "Documentation of cultural heritage using digital photogrammetry and laser scanning," *Journal of Cultural Heritage*, vol. 8, no. 4, pp. 423–427, 2007.

[154] A. Basu, S. Paul, S. Ghosh, S. Das, B. Chanda, C. Bhagvati and V. Snasel, "Digital Restoration of Cultural Heritage With Data-Driven Computing: A Survey," *IEEE Access*, vol. 11, pp. 53 939-53 977, 2023.

[155] Chan, E. R., Monteiro, M., Kellnhofer, P., Wu, J., & Wetzstein, G. (2022). Efficient Geometry-Aware 3D Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022) (pp. 16123–16133). IEEE.

[156] P. Grnarova, K. Y. Levy, A. Lucchi, N. Perraudin, I. Goodfellow, T. Hofmann and A. Krause, "A domain agnostic measure for monitoring and evaluating GANs," arXiv preprint arXiv:1811.05512, 2018.

[157] S. Jenni and P. Favaro, "On stabilizing generative adversarial training with noise," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 12145–12153.

[158] T. Karras, M. Aittala, S. Laine, E. Herva, and T. Aila, “Elucidating the design space of style-based generators,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 45, no. 1, pp. 87–104, Jan. 2023.