Generation of Textual Art Using Hybrid Learning Techniques

MCS

Author

Syeda Hafsa Ali

Registration Number

00000328725

Supervisor

Prof. Dr. Hammad Afzal

A thesis submitted to the faculty of Department of Computer Software Engineering, Military College of Signals (MCS), National University of Sciences and Technology (NUST), Rawalpindi in partial fulfilment of the requirement for the degree of MS in Computer Software Engineering.

(August 2023)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by **Ms Syeda Hafsa Ali** Registration No. **00000328725**, of **Military College of Signals** has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations/MS Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members and local evaluators of the scholar have also been incorporated in the said thesis.

Signature: Hamad Afzal

Name of Supervisor Dr. Hammad Afzal

Date:

Signature (HOD):

Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

Date: 7/9/23

Signature (Dean/Principal)

Date: 7/9/2023

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

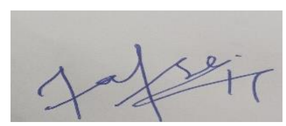
Declaration

I, Syeda Hafsa Ali declare that this thesis "Generation of Textual Art Using Hybrid Learning Techniques" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
5. I have acknowledged all main sources of help.
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Syeda Hafsa Ali , 00000328725.



Syeda Hafsa Ali,
00000328725 MSSE27

Dedication

“In the name of Allah, the most Beneficent, the most Merciful"

I dedicate this thesis to My supervisor: Dr. Hammad Afzal, Muneeba Daud, Jumana Noor, Majid Ali, my friends and teachers.

Abstract

Textual Art holds a significant role in preserving the traditions and heritage of various historical cultures. Particularly in the Eastern culture, it has served as an embellishment for homes and mosques, meticulously crafted by skilled artisans endowed with a sense of aesthetics. Modern endeavors have been directed towards digitizing this art form, yet the resources available online for textual art, especially in Arabic and Urdu styles, remain quite limited. Consequently, utilizing pretrained models yields unsatisfactory outcomes. The essence of geometric patterns and textures holds paramount importance in visual style, particularly within Islamic art, where these patterns bear profound symbolic and artistic significance. Our approach involves a fusion of deformable style transfer (DST), an optimization-centric technique that harmonizes the texture and geometry of a given content image to closely align with a chosen style image. This methodology is coupled with diffusion model inpainting, which is applied to content images originating from Calliar: an online dataset featuring handwritten Arabic calligraphy. Diverging from previous generative art methodologies, our approach presents aesthetically pleasing results, all the while preserving the integrity and structure of the Arabic script. Our methodology is showcased across a diverse spectrum of Arabic calligraphy writing styles, encompassing even the most prominent types, including Diwani ديواني, Thuluth ثلث, Kufi كوفي, Farisi فارسي, Naskh نسخ, and Rekaa رقعة.

Keywords: Neural style transfer, Arabic art generation, diffusion model inpainting, Calliar, Generative art

Acknowledgments

All praises to Allah for the strengths and His blessing in completing this thesis. I would like to convey my gratitude to my supervisor, Prof. Dr. Hammad Afzal, PhD, for his supervision and constant support. His priceless help of constructive comments and suggestions throughout the experimental and thesis works are major contributions to the success of this research. Also, I would thank my teacher Lt. Col. Khawir Mahmood for his support and guidance throughout and committee member; Assoc. Prof. Dr. Naima Iltaf for her support and knowledge regarding this topic. Lastly, I am highly thankful to my friends and family for their constant support throughout. I would like to thank them for their patience, cooperation and motivation in times of stress and hard work.

Contents

List of Figures

List of Tables

List of Abbreviations

CNN	***Convolutional Neural Network.***
NST	***Neural Style Transfer.***
GAN	***Generative adversarial network.***
DST	***Deformable Style Transfer.***
NLP	***Natural language processing.***
DM	***Diffusion Model***
LDM	***Latent Diffusion Model***
NBB	***Neural Best Buddies***
SIANN	***Space Invariant Artificial Neural Networks***
ReLU	***Rectified Linear Unit***
FoA	***Face of Art***
DuGAN	***Dual Generative Adversarial Network***
DCGAN	***Deep Convolutional Generative Adversarial Networks***

Chapter 1

1. Introduction

1.1 Overview

Textual art has played a significant role in various ancient heritages and cultures, including Chinese, Arabic, and postmodernist art utilising text-based elements, which has been embraced by numerous contemporary artists starting from the 1950s. Text-based artwork, characterised by the use of words and phrases, has been observed across various artistic mediums, encompassing painting, sculpture, lithography, screen printing, and applied art forms such as T-shirts and mugs. Projection mapping is also evident in the most recent iterations of modern art. The use of Arabic text art spans a significant duration, as it has been employed extensively for embellishing residences, mosques, and communal spaces throughout history. In contrast to conventional handwriting, text art introduces an artistic perspective by virtue of the creative liberty involved in shaping individual letters. There exist several distinct forms of text art, namely typography, lettering, and calligraphy. The Arabic calligraphy styles that have the highest level of popularity include Diwani, Thuluth, Kufi, Farisi, Naskh, and Rekaa. Regular handwriting can be regarded as a distinct form of calligraphy, characterised by the utilisation of specific styles, notably Naskh (نسخ) and Rekaa (رقعة).

Typically, the creation of calligraphy involves the physical expertise of individuals with a keen sense of aesthetics, supplemented by the assistance of computer tools. In recent years, there has been a significant endeavour to digitise this particular form of artistic expression using methods such as capturing photographs of embellished structures or rendering them digitally using electronic devices. However, it is important to note that the availability of internet materials for textual art, especially in Arabic and Urdu styles, remains very limited. Consequently, the availability of Urdu and Arabic textual art online is restricted, leading to subpar AI-generated art produced by pretrained generative models.

Despite the efforts made in other languages to create a universal solution for style recognition, it is not efficient or appropriate to Arabic due to its unique characteristics. The Arabic script exhibits variations in the form of its letters, which can either be isolated or joined based on their position within a word (initial, medial, or final). This characteristic poses a challenge for AI generative models in accurately predicting the sequential arrangement and aesthetic representation of Arabic letters. One additional challenge encountered in Arabic calligraphy is to the potential occurrence of characters overlapping vertically. Furthermore, the forms of letters are influenced by both the act of writing and the individual writer's stylistic choices, as well as the intended goal of the written text, such as decorative, documentary, or other objectives.

Typography can be considered as an interdisciplinary field that combines elements of technology and liberal arts. The emergence of neural style transfer and generative adversarial network has led to increased scholarly interest in Typography, specifically in the areas of text impact style transfer and new font style transfer. The objective of text effect style transfer is to generate text pictures that incorporate the visual style of other images, resulting in text effect images. However, the current approaches produce font pictures that are difficult to identify when dealing with more intricate text. In the context of font style transfer, the creation of novel styles relies on the modification of glyphs. The existing scholarly investigations in the field of textual art predominantly centre around the development of novel typographic styles and the generation of diverse textual manipulations.

The majority of contemporary methodologies [7,18,2,8,14,9,20,21,17] encompass a conceptualization of "style" that centres around the elements of colour and texture. Art historians and other scholars specialising in image formation, on the other hand, adopt a broader definition of style and consistently acknowledge the significance of shapes and geometric forms inside an artwork as essential components of its stylistic characteristics [13,6]. The recognition of an artist's individual style in many artistic mediums, such as painting (e.g., Picasso, Modigliani, El Greco), sculpture (e.g., Botero, Giacometti), and other kinds of media, is heavily influenced by the utilisation of shape and form. In order to achieve a stylized representation of a picture, style transfer algorithms employ a process wherein the visual content of one image is transformed to adopt the stylistic characteristics present in one or many other images. The current outcomes of style transfer algorithms have demonstrated remarkable achievements. However, it is worth noting that style transfer methods that do not explicitly incorporate geometry inside their style definition typically maintain the original geometry of the material in the resulting output. Consequently, the outcomes produced by these algorithms may be readily recognised as modified or "filtered" renditions of the original content image, rather than original images generated with the content image as a point of reference.

The proposed methodology involves utilising a text prompt and an image as input. The text prompt contains Arabic character that is intended to be generated as literary art, while the style image serves as a reference for design. The algorithm does a search within the Calliar dataset to identify Arabic script that meets the specified criteria. Calliar is an internet-based collection of data specifically designed for the study of Arabic calligraphy. This dataset encompasses a total of 2,500 sentences and has been meticulously annotated to facilitate stroke, character, word, and sentence-level prediction analysis. The collection comprises sketches that are represented as lists of (x, y) coordinates, denoting the trajectory of the pen during the stroke. The term "sketch" refers to a collection of strokes, with each stroke being an annotation denoting the type of stroke together with its corresponding (x, y) components. Every sketch is stored as a JSON file. The generation of an image from the Calliar dataset is accomplished by rendering the stored coordinates. To obtain the desired stylistic image, we employ web scraping techniques to gather Arabic calligraphy, calligraphic style art, mosaics, and paintings with similar stylistic attributes from online sources. In addition, an extensive assortment of calligraphy art was amassed, encompassing various styles such as Diwani, Thuluth, Kufi, Farsi, and numerous others. Deformable style transfer was employed in our study due to the inherent difficulty of learning style transfer in our issue domain. The wide range of unconstrained domains and styles cannot be reliably encompassed within a realistic training set. Hence, similar to previous style transfer approaches in this context, DST employs an optimization-driven technique that utilises a pre-trained and unchanging feature extractor obtained from a convolutional neural network (CNN) that was trained for ImageNet classification. The DST output data were integrated with steady diffusion techniques for the purpose of image inpainting. Stable Diffusion refers to a latent text-to-image diffusion model that has been pre-trained. This model exhibits the ability to generate photo-realistic images based on textual input. Additionally, it possesses the capability to do inpainting on images by utilising a mask. The Stable-Diffusion-Inpainting algorithm is initialised using the weight values from the Stable-Diffusion-v-1-2 algorithm. In the context of inpainting, the UNet architecture incorporates 5 more input channels. These channels consist of 4 encoded masked-image channels and 1 channel dedicated to the mask. The weights of these additional channels are initialised to zero upon the restoration of the non-inpainting checkpoint.

The complexity of generating original textual art surpasses the relatively straightforward process of transferring the style of one specific artwork onto the content of another. Recent research has focused on the acquisition of geometric style, employing either an explicit model of landmark constellations [25] or a deformation model that captures a particular style [23]. The utilisation of these methodologies necessitates a compilation of photographs adhering to the selected aesthetic, and is limited to a particular field of study (often focused on facial images, given their significance in societal and practical contexts). Diffusion models, Generative Adversarial Networks (GANs), and other Generative Models have been found to yield suboptimal outcomes when applied in isolation to Arabic art designs and language complexity. This might be attributed to the constraints imposed by inadequate training data and training biases. Therefore, their performance is suboptimal within the context of our scenario. However, in Chapter 4, we conduct a comparison between our results and those obtained from the approaches discussed earlier in the context of Text based Art (Arabic). Our findings indicate that our method yields significantly superior outcomes.

This study introduces a novel approach for creating legible Artistic Arabic Text Art designs by integrating the Calliar dataset, Stable diffusion for Image Inpainting, and DST, which incorporates geometry into one-shot, domain-agnostic style transfer. To the best of our knowledge, this is the first method proposed for generating such designs. The primary concept underlying the utilisation of DST involves seeking a seamless deformation, referred to as spatial warping, of the Calligraphy picture content. This deformation aims to align the image spatially with the reference style image. The process of deformation is facilitated by a predetermined selection of keypoints that aim to optimise the level of similarity between corresponding keypoints in both images. Following an initial step of approximately aligning the paired keypoints by a rigid rotation and scaling, a straightforward l_2 loss is employed to promote the warping of the output image in a manner that achieves spatial alignment of the keypoints. The regularisation of the deformation loss is accomplished by incorporating a penalty for total variation, which serves to mitigate the presence of artefacts resulting from significant deformations. This regularisation is then mixed with the conventional style and content loss terms.

The primary objective of this study is to produce Text Art images of superior quality, with a specific emphasis on Arabic script, while ensuring the preservation of its shape and structure.

Additionally, we aim to use shape and geometry as significant elements that contribute to the overall stylistic characteristics of the artwork. This is accomplished by the integration of a domain-independent geometric transformation of the content picture derived from an online dataset specifically designed for Arabic calligraphy, known as Calliar. This process is optimised in conjunction with conventional style transfer metrics, while also using a latent text-to-image diffusion model for the purpose of image inpainting. In order to provide a concise overview of the contributions made by this study,

We present Calli-Art, a novel hybrid learning system that aims to generate textual art by leveraging the capabilities of Calliar, the first online Arabic calligraphic dataset. Our approach integrates stable diffusion for image inpainting with deformable style transfer, which incorporates geometry into one-shot, domain-agnostic style transfer.

In this study, we present a novel approach to Arabic text-based art generation on the Calliar dataset by including geometry-aware style transfer. This is the first instance of such a demonstration in the field. In contrast to prior studies that exhibited subpar performance on Arabic datasets, our system yields visually enhanced outputs that are both intelligible and artistically appealing.

Our method is assessed on many Arabic calligraphy writing styles, encompassing the most prevalent calligraphy kinds such as Diwani ديواني, Thuluth ثلث, Kufi كوفي, Farisi فارسي, Naskh نسخ, and Rekaa رقعة, employing a user research for evaluation purposes. The findings of our study indicate that our framework has comparable performance to the most advanced models available, and it is capable of producing visually appealing artwork with a minimal dataset, while maintaining the structural integrity of Arabic script.

1.2 Motivation and Problem Statement

With the latest advancements in deep learning particularly with the development of generative adversarial networks multiple breakthroughs have been achieved in the area of typography and creative arts. Although there are multiple variations of Generative Models are utilized for producing textual effects or transforming different font styles, existing techniques frequently produce results highly dependent on training data and suffer training biases. when applied to a font from a different language, such techniques inevitably degrade certain visual artifacts of the text and in some cases the structural integrity of text is lost.

It is also quite hard task to develop a fully automated system for generating textual arts because flexibly changing font style and producing different effects while merging created fonts with background should all be considered.

Thus, the proposed study aims to fill the gap for textual arts generation driven by deep learning techniques..

1.3 Aims and Objectives

The following objectives are the focus of the research:

- Develop a framework driven by hybrid learning tehniques capable of novel textual art generation for different languages by learning a distribution of the style and content components of many different pieces of art, combining these components with text and create new pieces of textual art.
- Offer a range of supplementary resources and comparative experimental investigations on common datasets using state-of-the-art methodologies to illustrate the capability of our proposed paradigm.
- Carry out a study to assess the effectiveness and efficiency of our methodology by assessing the outcomes using objective evaluation.
- Demonstrate the potential of our proposed model by conducting a variety of ablation studies as well as comparative experimental studies on standard datasets with state-of-the-art approaches.

1.4 Research Contribution

To the finest of our knowledge, the methodology and framework presented in this thesis have not previously been used in the process of improving the quality of underwater images. The primary points of this thesis are:

- Implementing A hybrid learning framework for Artistis Textual Art generation using Calliar; first online Arabic calligraphic dataset, Stable diffusion for Image Inpainting and

Deformable Style Transfer which incorporate geometry into one-shot, domain-agnostic style transfer.

- Demonstrating, for the first time, geometry-aware style transfer for Arabic text based art generation on Calliar dataset. In contrast to previous works that performed poorly on Arabic datasets, authenticating our framework for giving readable, artistic and visually better results.
- Suggested method is evaluated by demonstrating it on a range of Arabic calligrapy writing styles including the most popular calligraphy types: Diwani, Thuluth, Kufi, Farisi, Naskh, and Rekaa and through a user study.

1.5 Thesis Organization

The thesis has been organized as follows:

- Chapter 2 provides a comprehensive survey of the relevant literature within the domain of Generative Art, with a specific focus on its manifestation as textual art in the context of visual form. Section 2.1 explores the foundational concepts intrinsic to generative arts within the visual arts domain. Section 2.2 gives briefing about generative arts models and techniques in the domain of visual arts. Section 2.3 provides a summary of the utilization of hybrid learning techniques for the generation of textual art. Importantly, this chapter undertakes a comparative analysis of the proposed system against the prevailing state-of-the-art methodologies in the field.
- Chapter 3 discusses materials and methodology used for proposed framework. It gives an overview regarding dataset collection, preprocessing, baseline models and discusses the proposed framework.
- Chapter 4 is results and discussion which presents results of the best baseline models applied and their limitations. It also provides an insight of the results improved by applying proposed framework for the particular dataset. The comparative analysis of previous studies is also presented which shows how art generation has improved with application of proposed framework.

- Chapter 5 is conclusion and future work which summarizes the research work, presents the limitations of the study and proposed framework with respect to calligraphy inspired textual art generation, it also suggests future direction in the domain.

Chapter 2

2. Related Work

This chapter conducts an extensive and insightful survey of pertinent literature within the domain of Generative Art. The chapter specifically directs its attention towards generative art in the form of textual art, set against the backdrop of its integration within visual expression. This section offers an in-depth examination of the available literature, focusing on multiple approaches to produce visual art that also prove valuable for generating textual art. The prevailing models for art generation primarily rely on techniques such as Style Transfer Methods, Generative Adversarial Networks, diffusion models, and data-driven or deep learning-based methods. The subsequent discourse categorizes and clusters diverse technologies and approaches employed in creating calligraphy-inspired textual art, based on their unique attributes.

2.1 Basic Concepts Associated With Generative Art

In this section, some basic concepts and models are explained which are closely related to generative models and in generative art design.

Neural Networks

A neural network is an artificial intelligence technique that imparts computers with the capacity to process data, drawing inspiration from the intricate workings of the human brain. Operating within the realm of machine learning, specifically under the banner of deep learning, it employs a layered arrangement of interconnected nodes, or neurons, mimicking the neural structure of the human brain. This architecture instigates the creation of an adaptive system, enabling computers to assimilate insights from errors and iteratively enhance their performance.

Deep Neural Networks

Deep neural networks, often referred to as deep learning networks, exhibit a configuration encompassing numerous concealed layers, wherein millions of artificial neurons are interlinked. Within this arrangement, connections between nodes are quantified as weights, signifying the influence one node holds over another, this weight can be positive, signifying excitation, or negative, indicating suppression. Nodes endowed with higher weight values wield heightened sway over their counterparts.

In theory, deep neural networks possess the capacity to establish mappings between any conceivable input and output formats. Nevertheless, this potency is coupled with an intensified demand for training.

Convolutional Neural Networks

A Convolutional Neural Network (CNN) represents a form of regulated feed-forward neural network that autonomously learns feature extraction through the optimization of filters, also referred to as kernels. CNNs manifest as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN) due to their shared-weight configuration of convolutional kernels or filters. These components traverse the input features, imparting translation-equivariant responses termed as feature maps. The architecture cconsists of an input layer, hidden layers, and an output layer. Within the hidden layers, one or more layers are dedicated to performing convolutions. This involves the computation of a dot product between the convolution kernel and the input matrix of the layer. The commonly employed activation function for this product is Rectified Linear Unit (ReLU). As the convolution kernel traverses the input matrix, a convolution operation is conducted, generating a feature map that subsequently contributes to the input of the ensuing layer. The sequence progresses to include other layers, such as pooling layers, fully connected layers, and normalization layers.

Neural Style Transfer

Style Transfer is a computational approach within the domain of computer vision and graphics, wherein a novel image is enerated by combining the content from one image with the stylistic attributes of another image. The underlying objective of style transfer pertains to the generation of an image that upholds the intrinsic content of its source while simultaneously infusing it with the visual demeanor found in a distinct reference image. Neural Style Transfer, emerges as an artificial construct grounded in Deep Neural Networks, yielding images imbued with a heightened perceptual quality and artistic essence. The architecture of this system leverages neural representations to segregate and subsequently reintegrate the content and style elements from disparate images, creating a neural algorithm for visually artistic imagery.

Generative Adverserial Networks

Generative Adversarial Networks (GANs) are designed for the training of generative models, using deep learning architectures within this framework. The GAN architecture was introduced by Ian Goodfellow et al. in 2014 titled "Generative Adversarial Networks," Subsequently, a standardized methodology referred to as Deep Convolutional Generative Adversarial Networks (DCGAN) was proposed by Alec Radford et al. in a paper published in 2015 titled "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." This formalization of DCGAN has led to the development of more robust and stable models within the GAN paradigm. GAN model architecture consists of two integral sub-models: a generator model, entrusted with the task of generating novel and credible instances within the problem domain, and a discriminator model, responsible for the classification of instances as either genuine - originating from the authentic domain, or synthetic - fabricated by the generator model. This structure creates a dynamic interplay between the generator and discriminator, characterized by adversarial training dynamics, wherein the generator strives to produce instances that can effectively deceive the discriminator, and the discriminator, in turn, tries to enhance its discernment capabilities in distinguishing real from generated instances.

2.2 Related Work

This section provide a detailed review of deep learning models which have been used for art generation.

Style Transfer Models

Previous approaches to style transfer in the early stages relied on manually designed features and algorithms [10–12,5]. However, Gatys et al. [7] introduced a novel method called Neural Style Transfer, which significantly advanced the state-of-the-art by utilising the features of a pretrained Convolutional Neural Network (CNN). In this approach, the style is represented using the Gramme matrix of features extracted from the shallow layers of the CNN, while the content is represented using feature

Optimization-based techniques have been shown to yield stylizations of high quality. However, their computational cost is a drawback due to the need for backpropagation at each iteration and the gradual modification of the image, typically at the pixel level, until the desired statistics are matched. To address this limitation, model-based neural approaches have been proposed. These approaches involve offline optimisation of a generative model, which enables the production of a stylized image with a single forward pass during testing. These methods can be categorised into two families, each with distinct tradeoffs compared to optimization-based style transfer. Certain methods prioritise speed and quality, resulting in rapid generation of excellent stylizations but limited to a predetermined set of styles. Other methods prioritise speed and allow for the transfer of arbitrary styles, but often produce lower quality outputs compared to optimization-based style transfer.

Until recently, the capabilities of style transfer methods were limited to transferring colour and texture, without the ability to transfer geometric style. However, in various studies unrelated to style transfer, geometric transformation has been successfully applied to images through automatic warping techniques. Early approaches involved predicting a set of global transformation parameters or a dense deformation field. Cole et al. (2019) introduced a method that allows for fine-grained local w

Several recent studies have endeavoured to integrate image warping and neural networks in order to acquire knowledge about both the textural and geometric styles of human portraits. For instance, CariGAN [19] employs a Generative Adversarial Network (GAN) to transform a photograph into a caricature. This is achieved by training one GAN to model geometric transformations using manually annotated facial landmarks, and another GAN to translate the typical non-geometric style appearances. Face of Art (FoA) [25], on the other hand, trains a neural network model to automatically detect 68 canonical facial landmarks in artistic portraits. These landmarks are then utilised to warp a photograph, aligning the facial geometry with that of an artistic portrait. Conversely, WarpGAN [23] introduces a warping module to the generator of a GAN and trains it as part of an end-to-end system. This is accomplished by optimising both the positions of keypoints and their displacements, using a dataset that comprises a caricature and photo pair for each individual.

In contrast, it should be noted that Deformable style transfer exhibits a broader applicability beyond human faces or any specific domain, and does not necessitate offline training on a meticulously curated dataset. Methodologically, the approach of Feature-wise Affine (FoA) entails the separate transfer of "texture" and "geometry," whereas Deformable Style Transfer (DST) performs their joint transfer. Conversely, WarpGAN also accomplishes joint texture and geometry transfer, but it necessitates the acquisition of paired examples in the face caricature domain for training its warping module, unlike DST which does not require such paired examples.

Two recent studies have employed CNN-based descriptors to establish connections between images, specifically focusing on non-human face domains. The first approach, known as Fully Convolutional Self-Similarity [16], utilises local self-similarity to identify matching keypoints across different instances within the same object class, thereby serving as a descriptor for dense semantic correspondence. The second method, called Neural Best-Buddies (NBB) [1], is a more general technique that aims to identify a set of sparse cross-domain correspondences by leveraging the hierarchical encoding of features obtained from pre-trained CNNs. In the context of deformable style transfer, NBB is utilised in conjunction with post-processing techniques.

Text and Image Style Transfer Methods

Text and image style transfer is a specific area of study within the broader field of neural style transfer. While image style transfer has been extensively researched over the years, the exploration of text image style transfer is a relatively recent development. In their work, Yang et al. (2017) were the first to investigate this problem and developed a patch-based model for migrating text effects. Another model proposed by Azadi et al. (2018) utilised depth-based techniques to stylize uppercase English letters given a small number of samples. However, this model can only generate images with a limited size of 64×64, and it is difficult to apply to other languagges or non englishh chharacters. Subsequently, Yang et al. [8] achieved a good transfer effect by manipulating glyph shape and style.

Score-Based Generative Models

Diffusion models (DMs), also known as Probabillistic diffusion models or score-based generative models, have attracted significant attention in recent times due to their stability and superior performance in image synthesis compared to Generative Adversarial Networks (GANs) [7, 24]. These models learn the data distribution by employing a Markov noising process, wherein a clean image I0 is subjected to noise at each step j, resulting in a set of noisy latent images Ij. The model is then trained to recover the original clean image I0 from the noisy latent images Ij in a backward process. DMs have demonstrated promising outcomes in various tasks such as unconditional image generation [7, 8, 25, 26], text-to-image generation [18–21], video generation [6], image inpainting [1, 2, 14, 16], image translation [15, 27, 30], and image editing [4, 5, 10].

Diffusion Based Text-Guided Inpainting Models

Taking advantage of the recent success of diffusion-based text-to-image generation models, an intuitive adaptation from a text-to-image generation to text-guided inpainting is to replace the pure random noise with the noisy background outside the mask region. However, this leads to strong artifacts, e.g., generating partial objects or inconsistent content in the background. To address this problem, GLIDE [16] generates a random mask and then provides the masked image and mask as additions to the diffusion model, which learns to utilize the information outside of the mask region. Blended diffusion [2] encourages the output to align with the text prompt using the CLIP score. Repaint [14] proposes to resample in each reverse step, but it doesn't support text input. PaintbyWord [3] pairs the large scale GAN with a full-text image retrieval network to enable multi-modal image editing. However, due to the structure of GAN, it cannot specifically modify the region given by the mask. TDANet [29] proposes a dual attention mechanism to exploit the text features about the masked region by comparing text with the corrupted image and its counterpart.

Studies Associated With Arabic Script

There Numerous studies have been conducted in the field of Arabic calligraphy. For instance, Koudja et al. (19) undertook a study involving the collection of 1,685 images, which were subsequently categorised into nine distinct classes representing different calligraphic styles. Allaf and Al-Hmouz (6) developed a system specifically designed for the recognition of various artistic Arabic calligraphy types. Their dataset consisted of a total of 267 images, with

AlSalamah et al. (31) conducted a study in which they obtained a dataset of 1000 calligraphy images from publicly available websites. The researchers manually annotated the dataset to extract letter images. Similarly, Bataineh et al. (8) gathered a dataset comprising 700 samples of various calligraphy styles, including Diwani, Rekaa, Kufi, Persian, Andalusi, Naskh, and Thuluth, from multiple sources. In another investigation, Bataineh et al. (7) assembled a dataset consisting of 14 Arabic degraded document images. Each image in this dataset represents a different challenge, such as low contrast, faded ink, multi-color elements, and presence of dirty spots.

Al-Hmouz (5) proposed an algorithm to recognise different types of Arabic calligraphy. The algorithm was evaluated using a combination of local and public datasets. The local dataset consisted of 18 images of sentences written by a skilled calligraphist. In this dataset, words were manually separated from the sentence images, and some words were rewritten independently due to difficulties in splitting them from the original sentence. The total number of samples in this dataset was 71 words. Additionally, a large computer-generated dataset called APTI (Arabic Printed Text Image) (33) was used. Khayyat et al. (21) collected 2653 images of historical Arabic manuscripts, which were categorised into 37 classes with six different handwriting styles.

In their study, Adam et al. (2) gathered a dataset consisting of 330 images of individual Arabic letters extracted from ancient manuscripts. This dataset was subsequently utilised for the purpose of classifying Arabic script styles through the segmentation of letters. In a separate publication (20), the authors introduced a computational tool named ACSR, which was designed to recognise and identify Arabic calligraphy styles from images that were captured.

The field of Arabic calligraphy and Arabic text art generation exhibits a significant research gap. While some research studies have focused on the classification of Arabic calligraphy styles, the availability of benchmark datasets remains limited. To the best of our knowledge, the Calliar dataset (1) stands as the sole benchmark for Arabic calligraphy, providing an online collection of handwritten Arabic text. This dataset encompasses 2500 annotated sentences and over 40,000 strokesgeneration.

2.3 Comparison With State-Of-The-Art Model

In this section, we present a comprehensive comparison between our proposed method and seven prominent generative art tools, namely DALL-E 2, Midjourney, Stable Diffusion, Imagen, Craiyon, Artbreeder, and NightCafe. The evaluation is based on key aspects encompassing creativity, control, accuracy, preservation of Arabic script integrity, input type versatility, and the ability to generate calligraphy-inspired designs.

Comparison of our method is based on following criteria:

Creativity: How creative and artistic are the images that the tool can generate?

Control: How much control does the user have over the output image?

Accuracy: How accurate are the images that the tool generates?

Arabic Script: Can the tool generate images that contain Arabic script?

Preservation capability: How well does the tool preserve the original image when generating a new image?

Input type: What types of input can the tool accept?

Ability to generate Calligraphy inspired designs: Can the tool generate images that are inspired by calligraphy

Generative Art Tools	Creativity	Control	Accuracy	Arabic Script Preservation	Input type	Calligraphy Inspired Designs
DALL-E 2	High	Medium	High	No	Text/Image	No
Midjourney	High	Low	Medium	No	Text/Image	No
Stable Diffusion	Medium	High	High	No	Text/Image	No
Imagen	High	Low	High	-	Text description	No
Craiyon	High	Low	Low	-	Text description	No
Artbreeder	Medium	Medium	Medium	-	Text description	No
NightCafe	High	Low	Low	-	Text description and user feedback	No
Proposed Method	**High**	**High**	**High**	**Yes**	**Text/Image**	**Yes**

Table 1: Comparison of Proposed Method with State of the Art Generative tools

Models for Art Generation	**Description**	**Content Preservation Ability**	**Limitations**
Neural Style Transfer	Create realistic and visually appealing results.	Can preserve the overall content of the image, but introduces artifacts.	Computationally expensive.
Image Inpainting	Fills in missing or corrupted parts of an image.	Can preserve the content of the image, but introduces distortions.	Less effective than Neural Style Transfer for creating realistic results.
DeepDream	Create psychedelic and abstract images.	Does not preserve the content of the image.	Useful for creative expression, but not suitable for image editing.
Generative Adversarial Networks (GANs)	Create realistic and diverse images.	Can preserve the content of the image, but introduce artifacts.	Computationally expensive to train.
Variational Autoencoders (VAEs)	Create realistic and smooth images.	Can preserve the content of the image, but introduce blurriness.	Generate blurry outputs for complex input images. Not suitable for extrapolation

Table 2: Comparitive Analysis of Automated Art Generating Models

Chapter 3

3. Methodology and Framework

This chapter introduces a novel hybrid learning approach that leverages Calliar, an extensive online benchmark dataset comprising over 2500 handwritten Arabic calligraphy sentences. The proposed approach aims to generate an initial content image by utilising a text prompt. Specifically, Calliar is employed to generate the content image, which is then combined with a reference style image and inputted into a deformable style transfer algorithm. The resulting output from the style transfer process, along with a mask of the image frame, is further processed through stable diffusion XL for image inpainting. This comprehensive pipeline ultimately yields a final Arabic Art Painting ot a portrait.

3.1 Calligraphy Datasets

We used two datasets in this study, one for text generation and and second for image referencing:

3.1.1 The Calliar dataset.

Calliar is an online dataset of handwritten Arabic calligraphy. It contains 2,500 annotated sentences, with more than 40,000 strokes. The dataset covers a wide range of calligraphic styles, such as Diwani, Thuluth, and Farisi. This diversity makes Calliar a unique resource for researchers studying Arabic calligraphy.

Calliar allows for capturing calligraphy at multiple levels, from strokes to characters to words to sentences. This makes it a valuable tool for a variety of tasks, such as classification and generation.

To help researchers efficiently use the dataset, Calliar uses a consistent representation. Each stroke is recorded as a list of (x, y) coordinates, representing the path of the pen. A sketch is defined as a list of strokes, where each stroke is annotated with its type.

Figure 1: Visualization of Calliar Dataset.

3.1.2 The Style Images Dataset.

DST requires content image and style images to be from the same domain, to get best results. We created a small dataset of 600 calligraphhy images by scrappingg the internet.

3.1.3 Preprocessing for Calli-Art

The Calliar Dataset is stored as a collection of sketches. Each sketch is a list of (x, y) coordinates, representing the path of the pen while drawing a stroke. A complete sketch is defined as a list of strokes, where each stroke is annotated with its type. Each sketch is saved as a JSON file, which can be easily converted to an image by drawing the coordinates. To avoid the problem of having images with different dimensions, the maximum size of each dimension (width or height) is pre-fixed to 600 pixels. To keep the aspect ratio of the images and the strokes, the larger dimension is scaled to 600 pixels and the smaller dimension is multiplied by the same factor. For example, if an image had a size of (400, 200), it would be rescaled to a size of (600, 300).

3.2 Proposed Method For Calligraphy Inspired Text Art Generation Using Hybrid Learning Techniques

With the latest advancements in deep learning particularly with the development of generative adversarial networks multiple breakthroughs have been achieved in the area of typography and creative arts. Although there are multiple variations of GANS and stable diffusion models for producing textual effects or transforming different font styles, existing techniques frequently produce results highly dependent on training data which largely comprise of Latin-based scripts. when applied to a font from a different language, such techniques inevitably degrade certain visual artifacts of the text and in some cases the structural integrity of text is lost. It is also quite hard task to develop a fully automated system for generating textual arts because flexibly changing font style and producing different effects while merging created fonts with background should all be considered. The core objective is to design a pipeline that can be a starting point towards automated generation of calligraphic art for cursive writing styles, which is readable and visually appealing ae well.

The present study introduces a proposed strategy that combines hybrid learning techniques for the creation of calligraphy-inspired text art.

As stated in the preceding section, the primary aim is to develop a pipeline capable of producing visually captivating and easily legible calligraphic artwork based on textual input. To achieve this, a fusion of machine learning and deep learning methodologies was employed for the development of calligraphy-inspired art..

3.2.1 Generation of Content Images from Calliar

The Calliar dataset is employed in this study due to its focus on Arabic calligraphy, which presents unique challenges due to the cursive nature of the script. In contrast, existing literature on sketch drawing and text-to-image methods predominantly focuses on English, as exemplified by GANwriting [18], Scrabble-GANs [14], DF-GANs [35], Be´zierSketch [11], sketchRNN [16], and DoodlerGAN [15].

The Calliar dataset was employed for the purpose of generating calligraphy sketches. This was achieved by visualising a collection of (x, y) coordinates that were stored in JSON files inside the dataset. The Calliar dataset comprises a total of 2,500 instances.

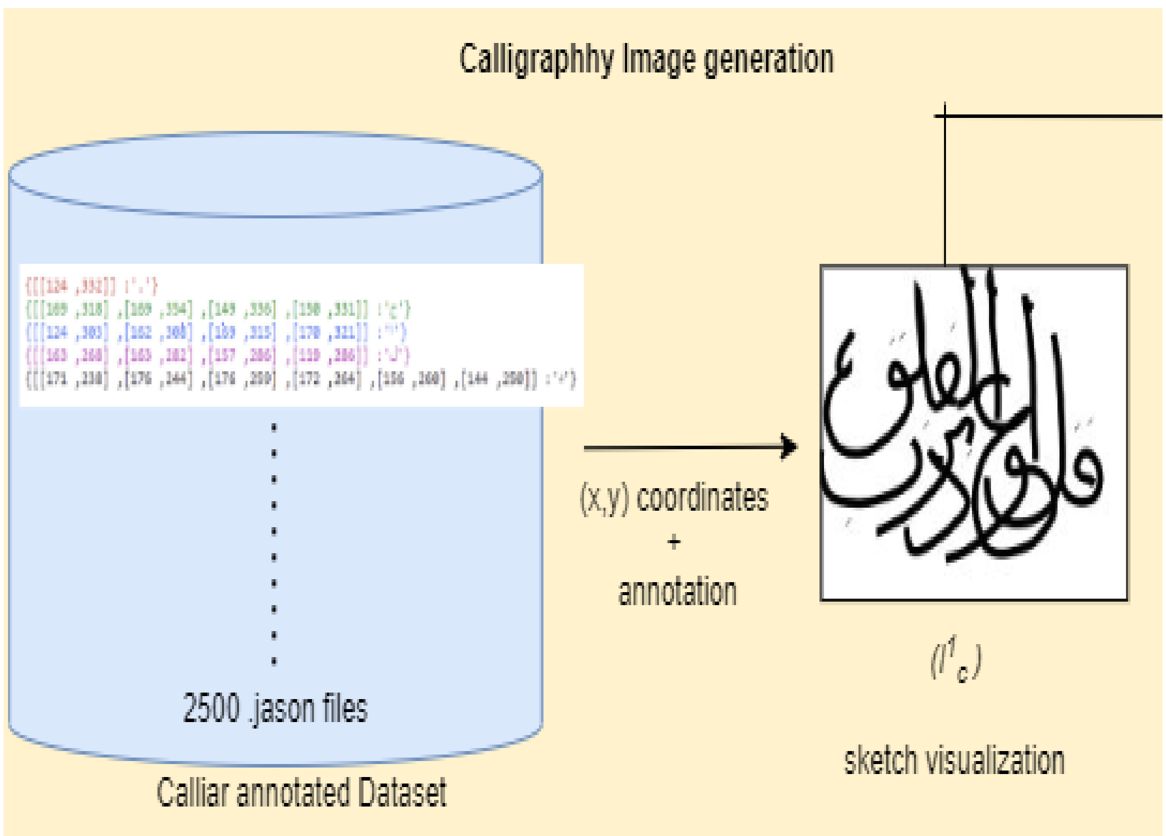


Figure 2: Calliar annotated dataset is visualized as Calligraphhy images

In this context, sentences are represented as lists of strokes, with each stroke serving as a representation of a single sentence.

The stroke sentence is composed with the inclusion of the (x, y) components. These components are then utilised to create an image through the process of coordinate drawing. Subsequently, this image is employed as the content image in a style transfer model, facilitating the development of calligraphic art.

3.2.2 Deformable Style Transfer in the Context of Calli-Art

In the context of a specific domain, the training set derived from that domain is utilised to identify significant and meaningful connections. This process involves employing facial landmark detection [25], training a data-driven detector for relevant points [16, 23], or having a user manually select matching points of two images while interacting with a style transfer tool. However, when it comes to generating calligraphic art with limited training data, it is not feasible to identify salient and meaningful connections through the training set from the domain. Therefore, in a one-shot, domain-independent scenario where such points are not readily available, we employ the NBB method utilised by DST. This method is a general approach for matching points between images.

The NBB algorithm is designed to identify a limited number of correspondences between two images, even if they belong to different domains or semantic categories. This is achieved by leveraging the hierarchical structure of deep features extracted from a pre-trained CNN. The deeper layers of the CNN capture high-level, semantically meaningful, and spatially invariant features, while the shallower layers encode low-level features like edges and colour information. The NBB algorithm starts by searching for pairs of correspondences that are mutual nearest neighbours, beginning from the deepest layer. These correspondences are then filtered based on their activation values and propagated through the hierarchy. This process helps to narrow down the search region at each level. Finally, the algorithm clusters the set of pixel-level correspondences into k spatial clusters and returns k keypoint pairs. The keypoint pairs returned by NBB, however, are often too noisy and not sufficiently spread out to use them as they are. To provide better guidance for geometric deformation, we closely follow DST[2] and use a modified NBB for Calligraphy inspired art. Specifically, we remove the final clustering step and return all pixel-level correspondences, usually on the order of hundreds of correspondence pairs. Then we use a greedy algorithm that selects a keypoint with the highest activation value (calculated by NBB) that is at least 10 pixels away from any already selected keypoint. We select up to 80 keypoint pairs and filter out keypoints with activation values smaller than 1. After the initial selection, we map the keypoints in the style image onto the content image by finding a similarity transformation that minimizes the squared distance between the two point clusters [24]. We then additionally clean up the selected keypoints by removing keypoints pairs that cross each other, to prevent a discontinuous warp field. We refer to the keypoints in the content image as the "source points" and the corresponding keypoints in the style image mapped onto the content image as the "target points." This process is illustrated in Figure 3.

The image deformation is defined by a collection of source keypoints Q = {q1,...,qk} and their corresponding 2D displacement vectors ε = {ε1,..., εk}. The destination coordinates for each source keypoint qi are determined by qi+εi. In accordance with previous work [23], we employ thinplate spline interpolation [4] to generate a dense flow field that maps the coordinates of an unwarped image U to a warped image W(U,ε). This is a closed-form procedure which finds parameters t,u,v that minimize $\sum_{i=1}^{k} \| f\varepsilon(q_i + \varepsilon_i) - q_i \|^2$ subject to a curvature constraint. With these parameters, we have the inverse mapping function

$$f_\varepsilon(q) = \sum_{i=1}^{k} t_i \phi(\|z - q_i - \varepsilon_i\|) + u^T z + v \quad (1)$$

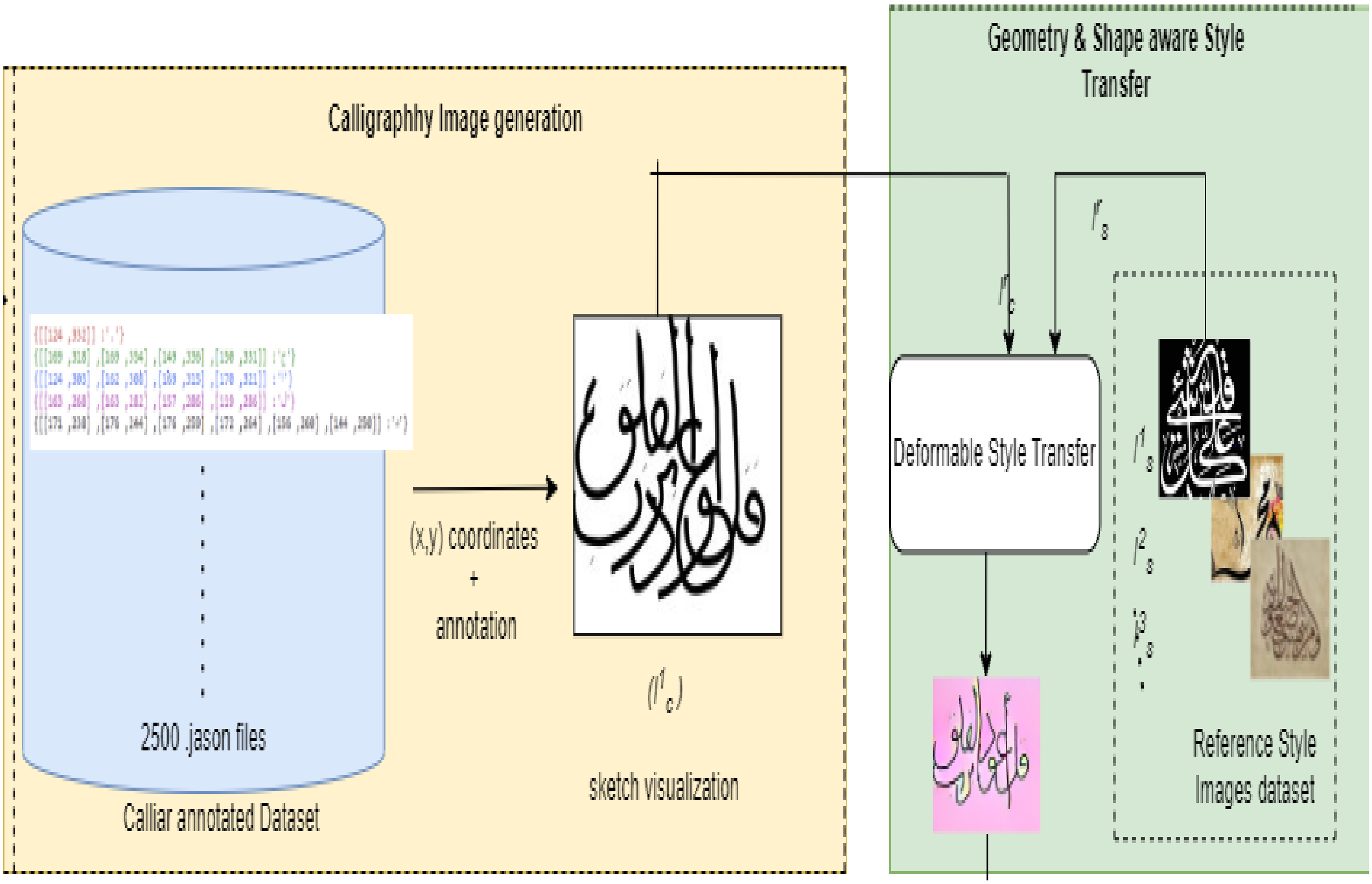


Figure 3: Stylization of Calliar Images using DST

3.2.3 Text-Guided Image Inpainting using LDM on Output image

Inpainting refers to the process of replacing masked regions in an image with new content. This is typically done to address image corruption or to remove undesired elements from the image. Text-Guided Image Inpainting is a technique that draws upon the principles of text-to-image generation to guide the inpainting process.

The achievement of state-of-the-art synthesis results on image data and other domains by diffusion models (DMs) is attributed to their decomposition of the image formation process into a sequential application of denoising autoencoders. However, the optimisation process for these powerful DMs is computationally intensive, often requiring hundreds of GPU days. Additionally, the inference stage is costly due to the need for sequential evaluations as these models typically operate directly in pixel space.

In contrast to pixel-based data models, our approach utilises Latent Diffusion Models that leverage pretrained autoencoders with limited computational resources, while still maintaining high quality and flexibility. Diffusion Models, as described in reference [79], are probabilistic models specifically designed to learn the distribution of data, p(x), by iteratively denoising a variable that follows a normal distribution. This process can be seen as learning the inverse of a fixed Markov Chain with a length of T. In the context of image synthesis, the most successful models, as discussed in references [15, 29, 70], employ a reweighted variant of the variational lower bound.

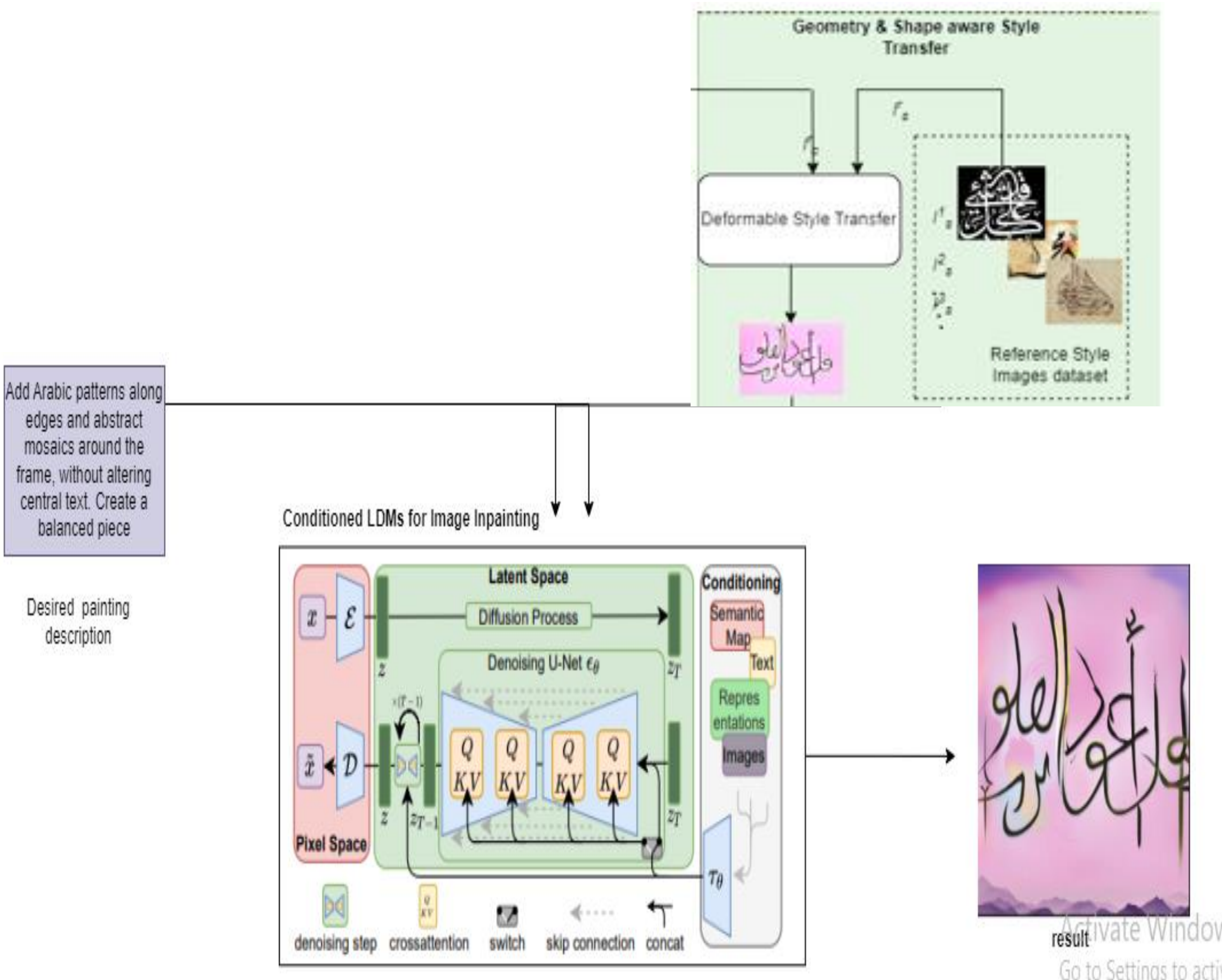


Figure 4: Image Inpainting using LDM on DST output

3.3 Experimental Analysis

By using hybrid learning techniques for Calli-Art generation we are able to achieve visually appealing art designs while keeping the structural integrity and readability of text. Using DST we alter the style of the calligraphy image generated from Calliar dataset without making it look like "filtered" versions of the original content, as they often do with standard style transfer methods. The results are further improved with LDMs by inpainting of output image. To the best of our knowledge, Calli-Art is the first effort towards automated generation of Arabic calligraphic art so we show the results of Calli-Art with state of the art generative art models

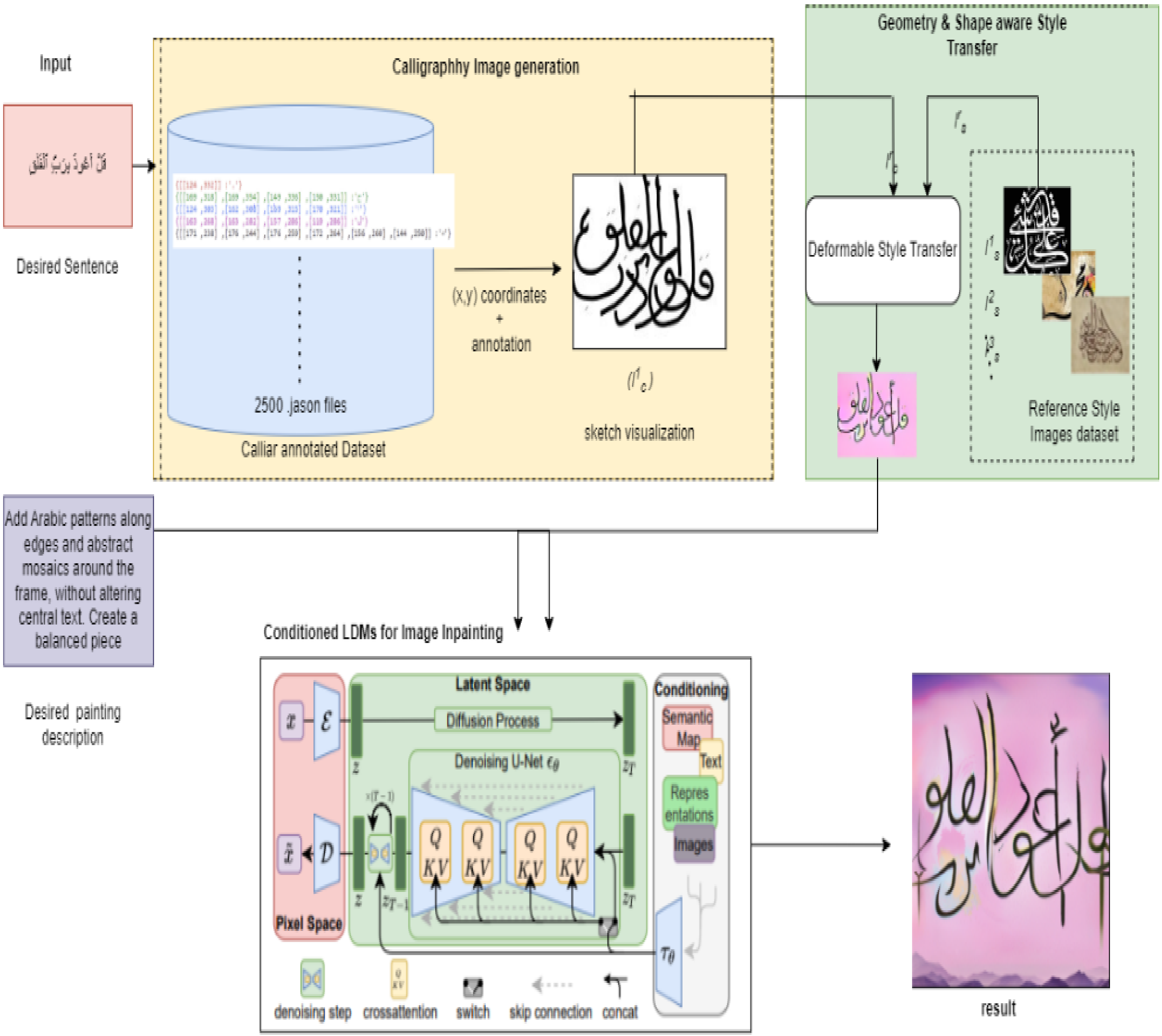


Figure 5: A detailed diagram of Calli-Art framework. It visualize input text prompt from calliar dataset. Use Deformable Style Transfer to create a stylized version of this image. Latent Diffusion Model improve this stylized image according to the text prompt provided by user by image inpainting and create an output image.

Chapter 4

4. Results and Discussion

To the best of our knowledge, Calli-Art is the first effort towards automated generation of Arabic calligraphic art so we show the results of Calli-Art with state of the art generative art models i,e DALL-E 2, Stable Diffusion, and Style 2 GAN in Figures 5 and 6. For a text prompt of Arabic text, we show the output of Calli-Art and the output of DALL-E 2, Stable Diffusion, and Style 2 GAN. To highlight the effect of the Call-Art pipeline, we also provide the comparison with output generated from Calliar and Deformable style transfer only. While DST and calliar produces readable Textual art, Calli-Art creates an aesthetically appealing textual art maintaining the readability and structural integrity of the language.

4.1 Comparison with Dall-E 2, Stable Diffusion and rasm (StyleGAN2)

To the best of our knowledge, Calli-Art is the first work towards automated calligraphic art generation in Arabic, other work in the domain of text art generation frequently produces results highly dependent on training data or are language dependent (mainly english), making them unfit for comparison. Since we consider Calli-Art as first step towards Arabic generative art, we draw the comparison of Calli-Art results with state of the art generative art models including, DALL-E 2[4] , Stable Diffusion [2] and rasm [5] in Figure 7. rasm is a StyleGAN2 model trained on a dataset of Arabic calligraphy and mosaic style images.

Calli-Art utilizes calliar dataset for visualizing Arabic text, DST for jointly optimizing the geometric and non-geometric stylization parameters, and LDMs for image inpainting. while DALL-E generates images from textual prompts by utilizing the VQ-VAE-2 to convert discrete textual descriptions into a continuous codebook space. This codebook representation is then employed to synthesize images that correspond to the given textual input. DALL-E uses CLIP, a large language model to encode text and images into a common embedding space.

When we compare Calli-Art and DALL-E in Figure 7, we provided a text description along with Arabic sentence for Arabic art generation.

The biggest drawback of rasm is that rasm uses StyleGan2 for Arabic Calligraphy generation. StyleGan2 requires a large amount of diverse and high-quality training samples to learn the necessary patterns and styles required for accurate generation. This results in visually appealing but rather incoherent and unreadable calligraphic art as sown in Fig 7. Moreover, rasm only generate generic Arabic calligraphy and is unable to generate calligraphic art in response to a prompt.

Stable diffusion XL uses diffusion models to generate high-quality images from text descriptions. They work by first training a diffusion model on a large dataset of images. The text description of image is encoded into a latent representation using a text encoder. This latent representation is used to initialize the diffusion model. Similar to DALL-E and rasm, Stable Diffusion XL faces training data biases and linguistic complexities for Arabic script. Since generative art models trained on languages with linear text structures, they struggle to capture the intricate connections between Arabic characters. This results in misalignment, distortion, and a lack of coherence in generated Arabic text art [Fig 10].

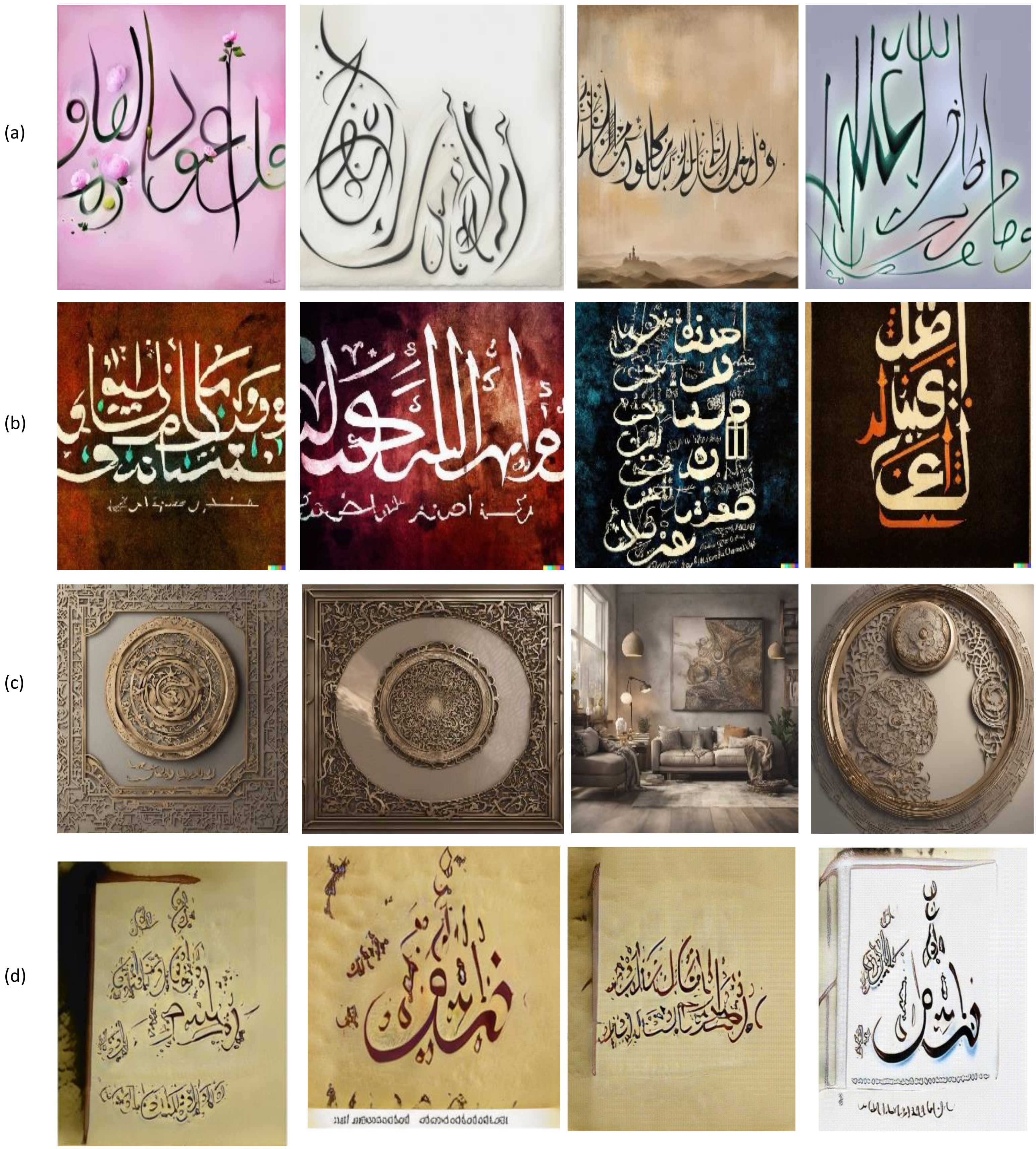


Figure 6: Calliggraphy Art results of Calli-Art framework on Calliar datasets. (a): Our method. (b): Dall E-2. (c): Stable Diffusion XL. (d): rasm (Style Gan2 usingg calligraphy images dataset)

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

Figure 7 Calli_Art results at each part. (a): Input sketch coordinates. (b): Calliar image visualization. (c): DST Output (d): Final Result (image inpainting by stable diffusion)

4.2 Qualitative Analysis of Proposed Approach via Human Evaluation

Quantitatively evaluating and comparing Calli-Art is challenging because it is difficult to objectively measure the aesthetic properties of style and visual quality. Additionally, there is an inherent trade-off between content preservation and stylization.

To address these challenges, we conducted a human evaluation study in a public university in Pakistan. We asked 25 students from the department of Fine Arts to evaluate the content preservation and stylization of Calli-Art, as well as two baseline generative art methods: DST and Latent Diffusion Model.

To measure content preservation, we asked the students the following question: "Do you think image A has the same sentence written as image B?", where A refers to the Calliar visualization image and B to the output of Calli-Art. The students could choose one of four answers: "Agree", "Agree, with slight difference", "Agree, with major difference", or "Disagree". We converted these answers to numerical scores (1 for "Disagree" and 4 for "Agree") and averaged the scores across all image pairs and users. This gave us a content score between 1 and 4 for each method.

To evaluate stylization, we presented the students with a pair of images, one from DST or Latent Diffusion Model and the other from Calli-Art, along with the Calliar image. The order of the first two images was randomized. We asked the students to choose which of the two output images better matches the Calliar image. The fraction of time a method was preferred in all comparisons gave us a style score between 0 and 1. A style score of 0.7 means that the method was preferred 70% of the time.

In total, there were 200 unique content comparisons and 150 unique stylization comparisons. 123 users participated in the content evaluation and 103 users participated in the stylization evaluation. The standard deviation of the content choice agreement was 0.79, and the standard deviation of the stylization choice agreement was 0.64.

These results suggest that Calli-Art is able to preserve content well while also stylizing the images. It is also more effective at stylization than the baseline methods.

Chapter 5

5. Conclusion and Future Work

5.1 Conclusion

This chapter, briefly summarizes the efforts, limitations, and recommendations for future studies. After explaining the conceptual approach, conducting experiments, and reviewing the findings the final observations and expositions are discussed. The goal of the project is to create a model that performs better than cutting-edge approaches. Our method, Calli-Art aims to generate visually appealing Arabic Calligraphic style art using a limited dataset while preserving the structural integrity of Arabic script. Notably, our work represents the first attempt to develop a fully automated framework for generating Arabic text art. Through our experiments, we observe promising outcomes in comparison to existing methods employed in various generative art models. The goal of this project is to create a model that performs better than cutting-edge approaches in generating visually appealing Arabic Calligraphic style art using a limited dataset while preserving the structural integrity of Arabic script. Our method, Calli-Art, achieves this by using a deep learning model that is trained on a dataset of Arabic text art. The model is able to generate new text art that is both visually appealing and structurally sound.

Our experiments show that Calli-Art outperforms existing methods in terms of both the quality of the generated art and the preservation of the structural integrity of the Arabic script. We believe that Calli-Art is a significant step forward in the development of automated frameworks for generating Arabic text art.

5.2 Limitations

By learning many samples, the methodology based on deep neural algorithms can lessen the effect of the complicated undersea environment. However, the dataset is crucial, as the existing dataset's coverage is still constrained. Deepest learning-based techniques, put more emphasis on improving full integration of the underwater imaging model. Therefore, maximizing the spatial features can bear good generalization performance. While Calli-Art achieves promising results, there are still some limitations to the approach. One limitation is that the model is trained on a limited dataset of Arabic text art. This means that the model may not be able to generate art that is representative of all styles of Arabic calligraphy.

Additionally, the model is not able to generate art that is completely original. The model can only generate art that is based on the data that it was trained on.

Another limitation of the approach is that it is computationally expensive to train the model. This is because the model requires a large dataset of Arabic text art and a powerful computing system.

5.3 Future Work

The network can be trained using perception-related loss function and introduce factors that are consistent with human interpretation, which will make the network more effective across wider range of scenarios.

Additionally, in order for the network to handle the loss of detailed data while taking into consideration speed, researchers can use the network's multi scale context features more frequently and specifying the step-wise reinforcement learning techniques while improving real-time performance and strengthen research on underwater video enhancement technology. There are a number of directions for future work on this project. One direction is to collect a larger dataset of Arabic text art. This would allow the model to generate art that is more representative of all styles of Arabic calligraphy. Additionally, it would allow the model to generate more original art.

Finally, it would be interesting to explore the use of Calli-Art for other applications, such as generating Arabic fonts or creating Arabic text animations.

6. References

1. Aberman, K., Liao, J., Shi, M., Lischinski, D., Chen, B., Cohen-Or, D.: Neural best-buddies: Sparse cross-domain correspondence. ACM Transactions on Graphics (TOG) **37**(4), 69 (2018)

2. Berger, G., Memisevic, R.: Incorporating long-range consistency in cnn-based texture generation. arXiv preprint arXiv:1606.01286 (2016)

3. Champandard, A.J.: Semantic style transfer and turning two-bit doodles into fine artworks. CoRR **abs/1603.01768** (2016), http://arxiv.org/abs/1603.01768

4. Cole, F., Belanger, D., Krishnan, D., Sarna, A., Mosseri, I., Freeman, W.T.: Synthesizing normalized faces from facial identity features. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)

5. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. pp. 341–346. ACM (2001)

6. Elkins, J.: Style. The Dictionary of art **29** (1996)

7. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)

8. Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)

9. Gu, S., Chen, C., Liao, J., Yuan, L.: Arbitrary style transfer with deep feature reshuffle. In: CVPR (2018)

10. Haeberli, P.: Paint by numbers: Abstract image representations. In: ACM SIGGRAPH computer graphics. vol. 24, pp. 207–214. ACM (1990)

11. Hertzmann, A.: Painterly rendering with curved brush strokes of multiple sizes. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques. pp. 453–460. ACM (1998)

12. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H.: Image analogies. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. pp. 327–340. ACM (2001)

13. Hofstadter, D.: Metamagical themas: Variations on a theme as the essence of imagination. Scientific American **247**(4) (1983)

14. Huang, X., Belongie, S.J.: Arbitrary style transfer in real-time with adaptive instance normalization. In: ICCV (2017)

15. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision. pp. 694–711. Springer (2016)

16. Kim, S., Min, D., Ham, B., Lin, S., Sohn, K.: Fcss: Fully convolutional self-similarity for dense semantic correspondence. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(3), 581–595 (March 2019). https://doi.org/10.1109/TPAMI.2018.2803169

17. Kolkin, N., Salavon, J., Shakhnarovich, G.: Style transfer by relaxed optimal transport and self-similarity. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

18. Li, C., Wand, M.: Combining markov random fields and convolutional neural networks for image synthesis. In: The IEEE Conference on Computer Vision and

19. Pattern Recognition (CVPR) (June 2016)

20. Li, W., Xiong, W., Liao, H., Huo, J., Gao, Y., Luo, J.: Carigan: Caricature generation through weakly paired adversarial learning. CoRR **abs/1811.00445** (2018), http://arxiv.org/abs/1811.00445

21. Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. arXiv preprint arXiv:1803.02077 (2018)

22. Sanakoyeu, A., Kotovenko, D., Lang, S., Ommer, B.: A style-aware content loss for real-time hd style transfer. In: CVPR (2018)

23. Selim, A., Elgharib, M., Doyle, L.: Painting style transfer for head portraits using convolutional neural networks. ACM Trans. Graph. **35**(4), 129:1–129:18 (Jul 2016). https://doi.org/10.1145/2897824.2925968, http://doi.acm.org/10.1145/2897824.2925968

24. Shi, Y., Deb, D., Jain, A.K.: Warpgan: Automatic caricature generation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

25. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **13**(4), 376–380 (April 1991). https://doi.org/10.1109/34.88573

26. Yaniv, J., Newman, Y., Shamir, A.: The face of art: Landmark detection and geometric style in portraits. ACM Trans. Graph.

27. **38**(4), 60:1–60:15 (Jul 2019). https://doi.org/10.1145/3306346.3322984, http://doi.acm.org/10.1145/3306346.3322984

28. Yeh, M.C., Tang, S., Bhattad, A., Forsyth, D.A.: Quantitative evaluation of style transfer. arXiv preprint arXiv:1804.00118 (2018)

29. Ahmad R, Naz S, Afzal MZ, Rashid SF, Liwicki M, Dengel A (2020) A deep learning based arabic script recognition system: benchmark on khat. Int Arab J Inf Technol 17(3):299–305

30. Al-Hmouz R (2020) Deep learning autoencoder approach: Automatic recognition of artistic arabic calligraphy types. Kuwait J Sci 47(3)

31. Allaf S, Al-Hmouz R (2016) Automatic recognition of artistic arabic calligraphy types. J King Abdulaziz Univ 27(1):3–17

32. Bataineh B, Abdullah SNHS, Omar K (2011) Arabic calligraphy recognition based on binarization methods and degraded images. In: 2011 international conference on pattern analysis and intelligence robotics, vol. 1. IEEE

33. Bataineh B, Abdullah SNHS, Omar K, Batayneh A (2013) Arabic-jawi scripts font recognition using first-order edge direction matrix. In: international multi-conference on artificial intelligence technology, pp 27–38. Springer

34. Chaker I, Benslimane R, Harti M (2011) Creation of al mabsout moroccan fonts format. In: 2011 colloquium in information science and technology, pp 3–4. IEEE

35. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258

36. Das A, Yang Y, Hospedales T, Xiang T, Song YZ (2020) Be´ziersketch: a generative model for scalable vector sketches. arXiv preprint arXiv:2007.02190

37. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp 248– 255. IEEE

38. Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartograph: Int J Geograph Inf Geovis 10(2):112–122

39. Fogel S, Averbuch-Elor H, Cohen S, Mazor S, Litman R (2020) Scrabblegan: semi-supervised varying length handwritten text generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4324–4333

40. Ge S, Goswami V, Zitnick CL, Parikh D (2020) Creative sketch generation. arXiv preprint arXiv:2011.10039

41. Ha D, Eck D (2017) A neural representation of sketch drawings. arXiv preprint arXiv:1704.03477

42. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

43. Kang L, Riba P, Wang Y, Rusin˜ol M, Forne´s A, Villegas M (2020) Ganwriting: content-conditioned generation of styled handwritten word images. In: European conference on computer vision, pp 273–289. Springer

44. Kaoudja Z, Kherfi ML, Khaldi B (2019) An efficient multipleclassifier system for arabic calligraphy style recognition. In: 2019 international conference on networking and advanced systems (ICNAS), pp 1–5. IEEE

45. Kaoudja Z, Kherfi ML, Khaldi B (2021) A new computational method for arabic calligraphy style representation and classification. Appl Sci 11(11):4852

46. Khayyat MM, Elrefaei LA (2020) A deep learning based prediction of arabic manuscripts handwriting style. Int Arab J Inf Technol 17(5):702–712