Heritage Science

RESEARCH Open Access

Generative innovations for paleography: enhancing character image synthesis through unconditional single image models

A. Aswathy[1*] and P. Uma Maheswari[1]

Abstract

Data scarcity in paleographic image datasets poses a significant challenge to researchers and scholars in the field. Unlike modern printed texts, historical manuscripts and documents are often scarce and fragile, making them difficult to digitize and create comprehensive datasets. Recently many innovations have been arrived on single image generative models for natural images but none of them are focused on paleographic character images and other handwritten datasets. In paleographic images like stone inscription characters, maintaining exact shape and structure of character is important unlike natural images. In this paper we propose an unconditional single image generative model, CharGAN for isolated paleographic character images. In the proposed system, augmented images are generated from a single image using generative adversarial networks, while maintaining their structure. Specifically, an external augmentation inducer is used to create higher-level augmentations in the generated images. In addition, the input to the generator is replaced with dynamic sampling from a Gaussian mixture model to make changes to the low-level features. From our experimental results, we infer that these two enhancements make single-image generative models suitable not only for natural images, but also for paleographic character images and other handwritten character datasets, the AHCD dataset, and EMNIST, where the global structure is important. Both the qualitative and quantitative results show that our approach is effective and superior in single-image generative tasks, particularly in isolated character image generation.

Keywords Generative adversarial networks, Single image generation, Isolated paleographic character image, Augmentation inducer, Data scarcity, Gaussian mixture distribution

Introduction

Paleography is the study of ancient and historical handwriting, and it is essential for understanding and interpreting of historical manuscripts and documents. Historical documents are often centuries old and susceptible to damage and decay over time [1]. The limited availability of paleographic images hampers efforts to study and analyze various aspects of ancient scripts and writing systems. Researchers may encounter difficulties in finding representative samples that cover a wide range of time periods, regions, and languages. Additionally, variations in preservation, image quality, and handwriting styles further compound the challenge of building comprehensive and diverse datasets. Data scarcity restricts the scope and depth of research in paleography, making it challenging to draw generalizable conclusions and accurately represent the complexities of historical writing traditions.

Synthetic data is a useful alternative to solve data scarcity in historical documents and other domains where large amount of real data is challenging or

*Correspondence:
A. Aswathy
aswathyachuth@gmail.com
[1] Department of Computer Science and Engineering, College of Engineering, Anna University, Guindy, Chennai 600025, Tamilnadu, India

expensive. Generative Adversarial Networks (GANs) [2] are a powerful framework with the capability of learning highly complex data distributions which has two competing models a generator and a discriminator. Generator generates synthetic data samples similar to a known data distribution and discriminator tries to differentiate the generated samples from real samples. However, training GAN requires a huge amount of data, which are not always available. All prior and current state-of-the-art models, such as DCGAN [3], WGAN [4], and Big-GAN [5], have the issue of training generative models that require large amounts of data.

Recently, many GAN models emerged that are trained using just one image. Among these models, SinGAN [6] is the benchmark model that contains a pyramid of fully convolutional GANs, each of which is responsible for learning the patch distribution at a different image scale. However, the layout of objects in the image is frequently distorted by this model's incoherent patch switching. It is therefore unsuitable for images where global structure is crucial. In addition to this, SinGAN leads to overfitting and does not produce many variances in images formed at higher scales. Other single-image generation models [7–11] that follow the same pattern as SinGAN suffer from similar problems.

We proposed guided and controllable character image generation for single-image models to address these shortcomings. We used an augmentation inducer to make high-level augmentations and used controllable latent noise input to have more influence over the characteristics and features of the generated character images. From the experimental studies, we realized that these modifications sound promising in overcoming data deficiency in paleographic datasets.

We experimented using segmented characters from paleographic datasets such as Tamil stone inscriptions and palm leaves, which are private and have a constrained amount of accessibility. The Fig. 1 shows the overall processing of a sample inscription image. The original image is enhanced using image processing techniques such as noise removal, smoothing, grayscale conversion, and binarization. Each character was extracted from an improved image using the bounding-box approach. We found that many segmented characters were isolated. We used single-image GANs because the other few shot GANs were insufficient under these conditions. In ancient written documents, only minor differences exist between similar characters. To identify the different features in these datasets, it is crucial to maintain their exact shape and structure. From the literature reviewed, we understand that this problem is not restricted to paleographic images and is common in Indian language datasets and other complex character datasets, such as Arabic and Chinese, Korean, Japanese and ancient Egyptian dataset [12].

The summary of contributions of this paper are as follows:

- A single image generative model (CharGAN) for paleographic character images is introduced in proposed system. This method differs from typical GAN methods, which often require large training datasets. A modified version of the cutting-edge SinGAN architecture was employed with a guided generation and complex prior as input.

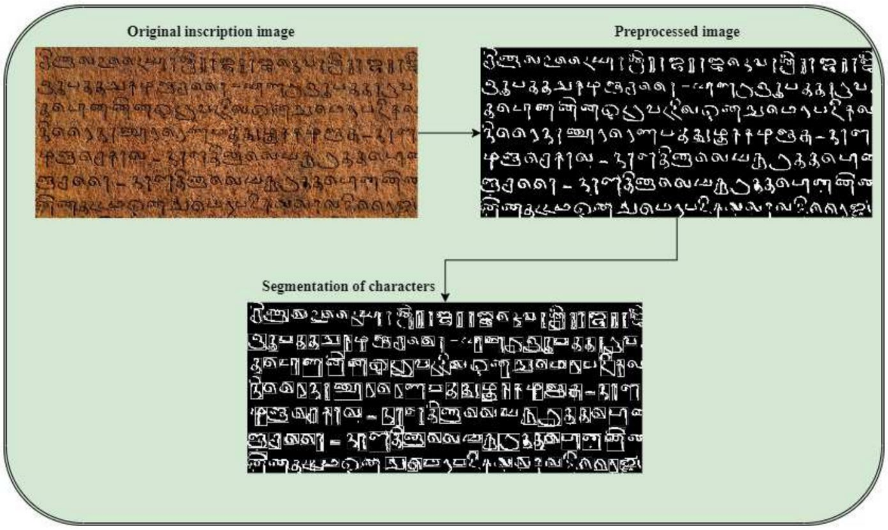

Fig. 1 Overall processing of a paleographic image

- Unlike previous single-image models, the proposed single-image unconditional GAN generates diverse images (without distorting objects) by utilizing an augmentation inducer with a specific focus on character datasets while maintaining its global structure.
- CharGAN unveils a Gaussian mixture latent code as the generator's input for single-image generative models, giving the network more editable and changeable properties, and delivering more realistic and diversified outcomes.
- We compare CharGAN to various generative models and benchmark them for GAN performance when just an image is available for training.

Research aim

The aim of this study was to develop and evaluate an unconditional single-image generative model tailored specifically for paleographic character images and other handwritten character datasets, with a focus on maintaining the precise shapes and structures of characters. Character datasets, particularly old paleographic datasets, differ from natural images or textures in that there is a high degree of character similarity, and we want to preserve exact structure and shape of each character. Consequently, when augmentation is applied to these datasets, the exact structure of the original characters should be preserved. This study addresses the challenge of data scarcity in paleographic image datasets by proposing a novel approach that leverages generative adversarial networks (GANs) and external augmentation techniques to create comprehensive datasets of augmented images. Additionally, this research aims to investigate the effectiveness and superiority of the proposed approach in comparison to existing methods, particularly in the context of isolated archaic character image generation.

Related work

Single image GANs: Several recent studies have focused on single-image GANs. InGAN [13] was the first to concentrate on the synthesis of natural images, which trains on a single input image and learns the internal distribution of the patches. Once trained on the input image, it can remap the input to any size or shape, while preserving the same internal patch distribution. However, it is a conditional generative model with no semantic understanding of the input image. Tamar et al. introduced SinGAN [6], an unconditional generative model to overcome the drawbacks of InGAN, which was trained on a single natural image. In addition to image generation, this method is applicable to a variety of applications such as harmonization, super-resolution, and animation. This is achieved using a pyramid of fully convolutional lightweight GANs, each of which is accountable for understanding the distribution of patches at different scales. The outstanding performance of SinGAN demonstrates the feasibility of internal learning on the generation task. ConSinGAN [7] is an extended SinGAN that rescales for multistage training and training several stages concurrently, allowing the model to be smaller and the training to be more efficient. ExSinGAN [8] trains three modular GANs to describe the distributions of the structure, semantics, and texture to produce a comprehensible generative model. SIV-GAN [9] is a model that learns from a single image or video to generate new plausible compositions of a given scene with varied content and layout. The GPNN [10] emulates SinGAN, but replaces the GAN unit with the classical patch nearest-neighbour module. Similar to GPNN, GPDM [11], which is based on patch nearest neighbours, directly minimizes the Sliced Wasserstein Distance between the output and target patch distribution, whereas GPNN approximately minimizes bidirectional similarity.However, because nearest-neighbor algorithms have a very limited scope of generalization, they can only be used for image generation processes in which it is possible to copy some parts of the input. SinFusion [14] is the first diffusion model for a unified framework that handles a wide range of single-image- and single-video-generation tasks. They employed denoising diffusion probabilistic models (DDPM) [15] for single-image generation tasks, which are trained on multiple random crops from the input image. SinDiffusion [16] is a diffusion-based system that extracts patch statistics from a single natural image. The model employed single-scale training and a network with patch-level receptive fields to generate the images. SinDDM [17], a denoising diffusion model for a single image, utilizes a multi-scale diffusion process and a fully convolutional denoiser to learn the internal statistics of the training image. By driving a reverse diffusion process, SinDDM can generate diverse high-quality samples with arbitrary dimensions in a coarse-to-fine manner.

Although existing methods produce realistic images from a single image, they all have severe flaws like extreme unpredictability and uncontrollability. They are unable to generate satisfactory images when the global structure needs to be maintained. In comparison, CharGAN can create random samples with accurate structures and varying looks from a single image based on user expectations.

Method

We describe the proposed system in detail in this section. The proposed model has a pyramid structure similar to that of SinGAN; however, it has two important architectural differences. An Augmentation inducer is employed in the generators of the GAN model to generate additional variability in the generated images. The second major change is in the case of latent code introduced at different GANs. In our model, choose one of the n Gaussian components at random and use the reparameterization trick [18] to sample from the chosen Gaussian distribution. The complex prior makes changes to low-level details, whereas the augmentation inducer is used to apply higher-level augmentations to the generated image.

The overall architecture of the proposed system is depicted in Fig. 2. CharGAN comprises a pyramid of GANs $\{GAN_0, GAN_1, ..., GAN_n\}$. GAN_i is composed of an adversarially trained generator G_i and a discriminator D_i. The generators $G_0, ..., G_n$ have two inputs: dynamic Gaussian mixture latent codes $\{L_0, L_1, ..., L_n\}$, and augmentation inducer A. Each generator G_n is responsible for generating realistic image samples based on the patch distribution in the corresponding image I_n and the augmentation inducer A.

The training process begins at the coarsest scale of the image pyramid and progresses through all generators until it reaches the finest scale, with noise and augmentation inducer introduced at each stage. At the lowest scale, noise L_n is fed into the generator, and at higher levels, an upsampled image from the previous

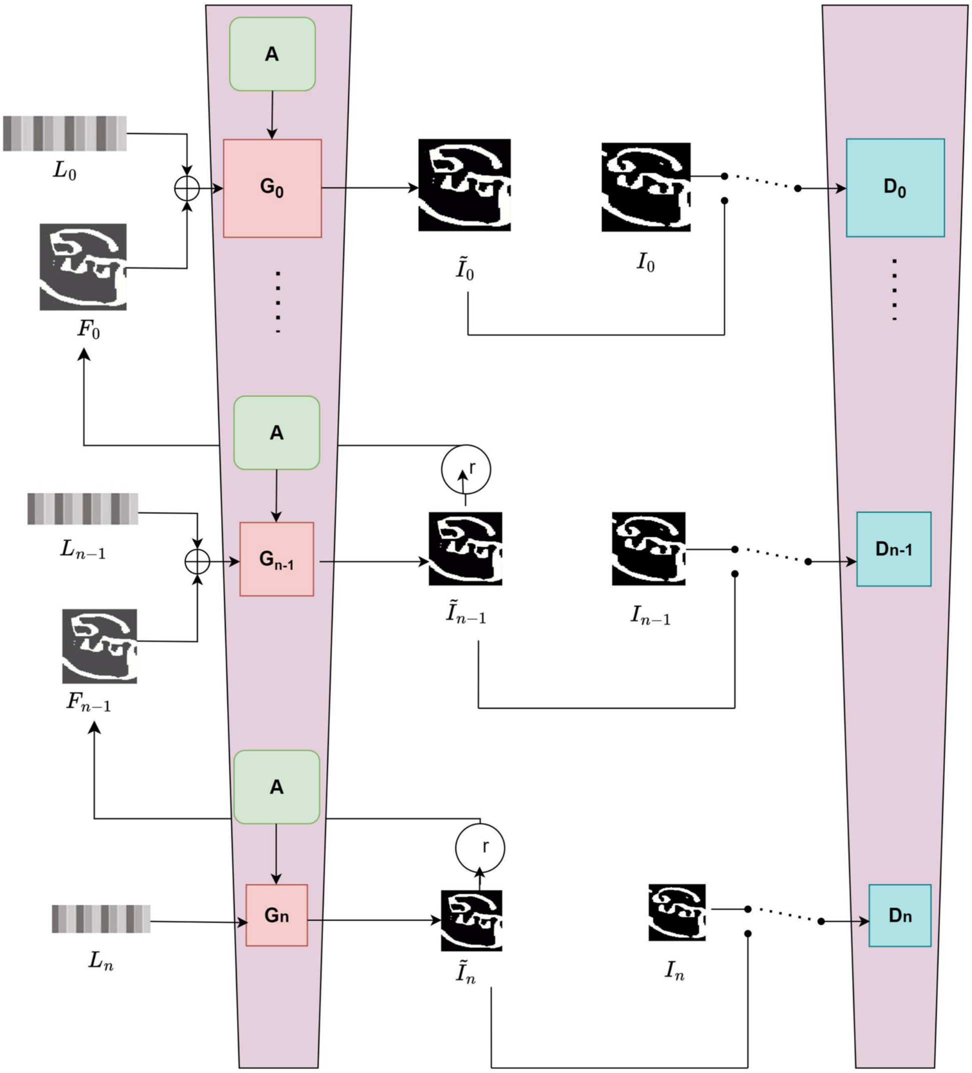

Fig. 2 Architecture of proposed single image generative model

scale, F_n, is also fed into the generator.The generator generates fake image I_n as :

$$\tilde{I}_n = G_n(L_n, F_n, A) \tag{1}$$

For each I_n with the corresponding GAN_n, G_n learns to map the randomly sampled noise from the Gaussian mixture model and the upsampled output of G_{n+1} into a fake sample. At finer scales, each generator G_n adds information that the previous scales did not generate. All generators have the similar architecture and receptive fields ; only the input and output sizes change.

Architecture of generator in detail is shown in Fig. 3. At each scale n, the image from the previous scale F_n is up-sampled and combined with the input noise map L_n. The resulting data is then processed through five convolutional layers, the outputs of which form a residual image. This residual image is added back to F_n to produce an output fake image, represented as $\tilde{I}_n$.

Augmentation inducer

To increase the diversity of generated images, we introduce a novel unit in CharGAN called augmentation inducer A, which is a learned geometric transformations applied to a single original image I_n. These learned transformations can be used to give higher-level information of the input image to the generator. One important consideration here is that we used augmentations that preserve the exact structure of the characters rather than augmentations like cropping and flipping.

Fig. 4 shows details of Augmentation inducer. I_A is a set of images created after performing geometric changes to a single input image, such as scaling, padding, and random distortions.

$$I_A = GT(I_n) \tag{2}$$

The augmentation inducer uses I_A, a group of images formed after making geometric modifications to a single input image, such as scaling, padding, and random distortions and outputs a weight w_n and bias b_n. These learned parameters are referred to as augmentation inducer A which is given to the generators at all scales to govern the augmentation of the original image. w_n is multiplied by GAN's original flow, and b_n is added. The network learns the image style by performing ordinary convolution followed by a cascade of fractional-strided convolutions (FS Conv) [19]. Before each FS Conv layer, we applied batch normalization (BN) [20] and leaky ReLU (LReLU).

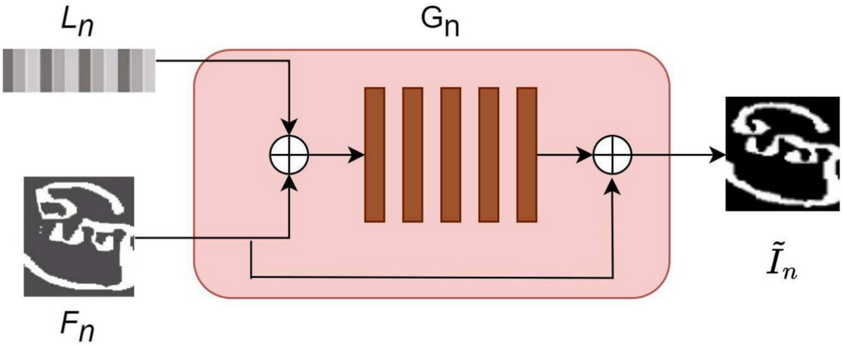

Fig. 3 Generator architecture

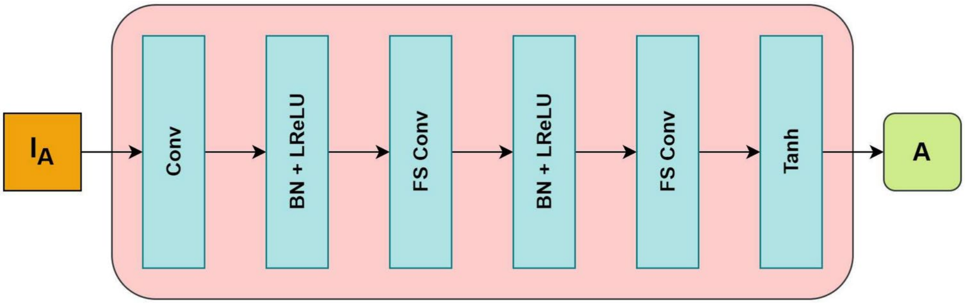

Fig. 4 Generation of augmentation inducer

Complex prior to generator

The most typical input to a GAN generator is random noise with a specific distribution. In CharGAN, we used more complicated prior-dynamic sampling from a Gaussian mixture model as the generator's input, which helps to produce more realistic and diversified images. Our technique can be considered an attempt to alter the latent space to gather samples in high-probability areas of the latent space. We obtained the latent code by sampling from n Gaussian mixture distributions instead of simply one Gaussian distribution. During the process of GAN training, our primary goal is to establish a mapping from a basic latent distribution p_L to a more intricate data distribution (Eq. 3).

$$P_L(L) = \sum_{i=1}^{M} \phi_i g(L|\mu_i, \Sigma_i) \tag{3}$$

where $g(L|\mu_i, \Sigma_i)$ represents the probability of the sample L in the normal distribution,(μ_i,Σ_i) is the mean and variance of Gaussian distribution n_i , weight of each Gaussian component is denoted by ϕ_i .

The neural network's backpropagation process relies on derivative computations during training. If we directly sample from discrete Gaussian mixture distributions, it would lead to gradient vanishing and hinder model updates. To overcome this challenge, we employ the reparameterization trick [18]. This involves randomly selecting one of the n Gaussian mixture components, denoted as ϵ, with a mean of 0 and a standard deviation of 1. We then adjust and flatten it to obtain the mixed random noise.

$$L = \mu_i + \sigma_i \epsilon \tag{4}$$

where ϵ is sampled from a standard normal distribution. Partial derivatives μ_i and σ_i are acquired to involve in back propagation. Determining L for propagation parameters optimizes μ and σ. As the weight in Eq. (4) remains non-trainable within the neural network, we set equal weights $\emptyset_i = \frac{1}{n}$. We use a dynamic factor λ to improve diversity and statistically modify the latent codes. This method allows for dynamic regulation of both μ_i and σ_i.

$$L = (1 - \lambda)\mu_i + \lambda \sigma_i \epsilon \tag{5}$$

Several works [21–23] have proposed using a mixture model for the latent space in the context of variational inference. Renzede et al. [24] and Kingma et al. [25] proposed 'normalizing flows' to generate a complex distribution by transforming the latent probability density through a series of invertible mappings. To the best of our knowledge, no such methodologies exist in the domain of single-image GANs. This will help to solve the overfitting problem in single-image GANs and to make feature-level changes in the generated images.

Training details

The essential parameters for network operation are detailed here to aid in understanding some settings in the following experiments. Initially, we preprocessed the images entering the network, primarily by resizing them such that the maximum dimension did not exceed 250 pixels and the minimum dimension was no less than 25 pixels.

We utilized a stage-wise training strategy, initializing the generator's weights and discriminator to those from the previously trained scale at each stage. We set the number of scales to eight, each undergoing 3000 iterations. The initial learning rate of the Adam optimizer [26] was set to 0.0002, decaying by a factor of 1/10 after every 1500 iterations. The momentum parameters β_1 and β_2 were set to 0.3 and 0.99, respectively. Batch Normalization (BN) was used to reduce overfitting during the training of both the Generator and Discriminator. The learning rate for both the Generator and Discriminator was 0.0002. Additionally, the LeakyReLU (LReLU) activation function was applied to prevent overfitting by adjusting the negative slope of LReLU when the BN was insufficient. For any scale, we set the LReLU to 0.2.

Loss functions

Each generator G_n is attached to a Markovian discriminator D_n, which determines whether the overlapping patches of its input are real or fake [27, 28]. In addition to the standard adversarial loss denoted as L_0, we apply the WGAN-GP loss [29] denoted as L_W to stabilize the training process of D_n. WGAN-GP loss can be expressed as:

$$L_W = \left(\left\|\nabla_{\bar{I}_n} D\left(\bar{I}_n\right)\right\| - 1\right)^2 \tag{6}$$

The discriminators' overall loss function can be written as:

$$L_D = L_0(G_n, D_n) + \lambda L_W(D_n) \tag{7}$$

where λ is the gradient penalty coefficient. In addition to classical adversarial loss, we use mean squared error (MSE) as the loss function for G_n.MSE given as:

$$L_1 = \left\|\tilde{I_n} - I_n\right\|^2 \tag{8}$$

The generators' final loss function is as follows:

$$L_G = L_0(G_n, D_n) + \alpha L_1(G_n) \tag{9}$$

where α is adam hyperparameter.

Results and discussion

We evaluated CharGAN both qualitatively and quantitatively on historical character datasets such as Tamil stone inscriptions and Palm leaf manuscripts. In addition to ancient character datasets, the model evaluated on domains of handwritten characters to determine its effectiveness. Specifically, the AHCD database of handwritten Arabic characters [30] and standard EMNIST [31] letters were used in this study. We compared the performance of CharGAN with other single-image models on random image generation, in both qualitative and quantitative terms.

Dataset

The stone inscription images used in this study were sourced from prominent Chozha Temples in Tamil Nadu, including the Brihadeeswarar Temple in Tanjore, Gangai Konda Chozhapuram, and Iravadeeswar Temple in Darasuram, following approval from the Archaeological Survey of India (ASI). Dating back to the 11th century, these images were captured using a high-resolution 77D 22.4 MP DSLR camera with an EF-S: 18 to 55 mm lens. Approximately 3000 images were initially captured and post-processed, which involved fine-tuning and size reduction of the raw images. A series of preprocessing steps were then applied to the images, including grayscale conversion, noise removal, morphological operations such as erosion and dilation, opening and closing operations, binarization, and normalization. The images were normalized to dimensions of 2400 x 1800 pixels. Each script image contained a minimum of 40 characters and a maximum of 200. Following character segmentation, 76,246 individual character images were obtained.

Palm-leaf manuscripts were housed within the Tamil Nadu Oriental Library. Permission obtained from the Department of Archaeology, Tamil Nadu Circle, to acquire digital copies of these manuscripts.Preprocessing is carried out using a procedure similar to that employed for the stone inscriptions. A total of 800 manuscripts were processed, resulting in approximately 10,000 characters being obtained.

One hundred classes were identified in the stone inscription datasets and approximately 40 classes were identified in the palm leaf datasets. There were 32 classes in total among these characters, with a count of less than ten. We trained CharGAN on these isolated characters and created characters that could be used for further recognition and translation of these ancient characters into contemporary Tamil. In addition to this, CharGAN trained on the AHCD database of 28 classes of isolated handwritten Arabic characters collected by and handwritten English letters from the EMNIST dataset. We chose the AHCD dataset because of the complicated character structure of Arabic characters and the EMNIST dataset as a base.

Qualitative results

To demonstrate the efficiency of our model, we conducted a comprehensive comparison that included both single-image Generative Adversarial Networks (GANs) and diffusion models. By evaluating our method against these two prominent types of image-generation techniques, we aim to provide a thorough validation of our model's performance. The qualitative results of

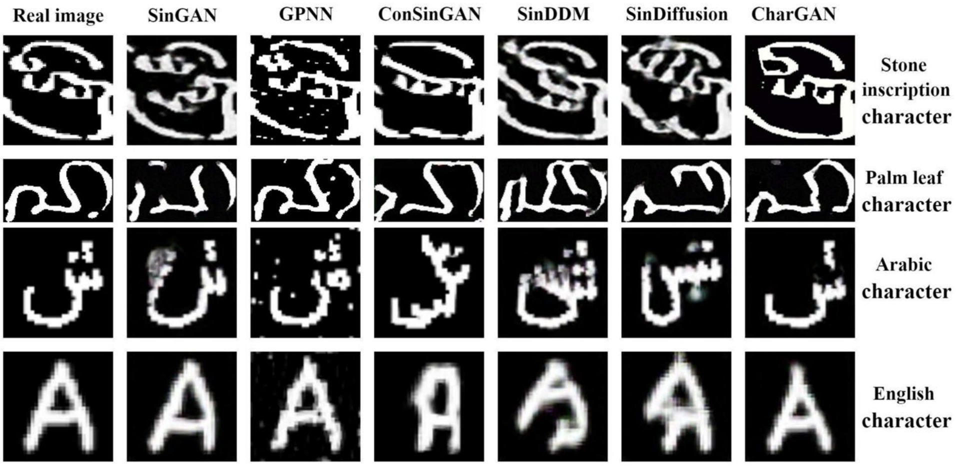

Fig. 5 Randomly generated samples from SinGAN, GPNN, ConSinGAN, SinDDM, SinDiffusion and CharGAN for four different datasets

sample images from four datasets generated by CharGAN and other single-image models using GAN (SinGAN [6], GPNN [10], and ConSinGAN [7]) and single image diffusion models (SinDDM [17], SinDiffusion [16]) are shown in Fig. 5. The leftmost column is the real images, and to the right, in turn, are the results of SinGAN, GPNN, ConSinGAN,SinDDM,SinDiffusion and CharGAN, respectively. The figure shows that the existing single-image models fail to preserve the exact structure of characters or generate diverse images, particularly for complicated characters. Our methodology generates images with visible variations in appearance as shape and posture. In the case of the EMNIST dataset, almost all GAN models produce acceptable images, although other datasets do not. When we examined different characters in similar datasets, the same pattern occurred. Interestingly we realized that in the case of character datasets, models with a pyramidal structure (SinGAN, GPNN, and CharGAN) outperform others such as ConSinGAN.When it comes to single image diffusion models, our results clearly indicate that existing models are not suitable for character generation. These models are effective at producing diverse images, but they often fail to preserve the global structure. This limitation highlights a significant research opportunity in developing diffusion models specifically tailored for single image character generation. Additional samples generated by CharGAN are shown in the Fig. 6. The images created by CharGAN appear to be more plausible and reasonable, which proves the superiority of the model.

Quantitative results

Table 1 shows a quantitative comparison between CharGAN and state-of-the-art single-image models. We analyse the quantitative performance of single image

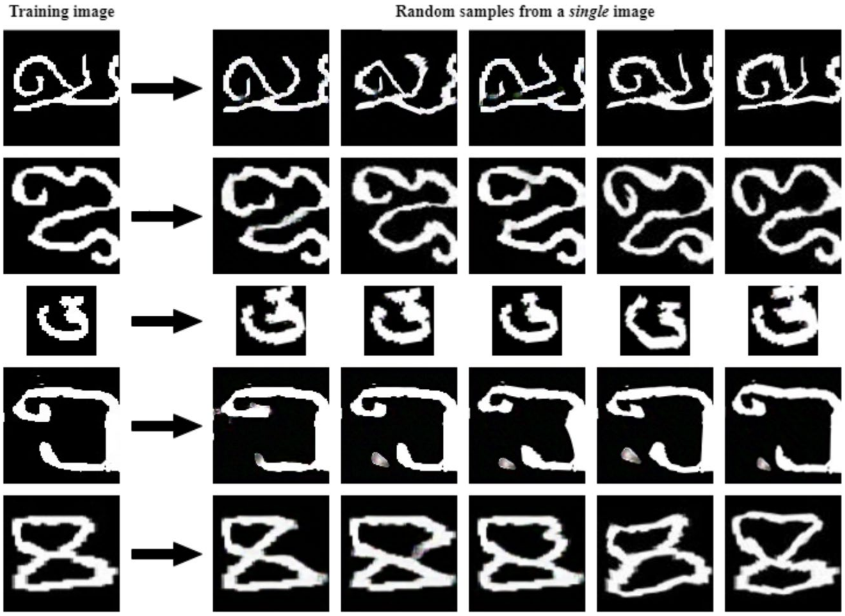

Fig. 6 Randomly generated samples using CharGAN

Table 1 Quantitative results of different Single image generative models

Method	Stone		Palm leaf		AHCD		EMNIST	
	SIFID	LPIPS	SIFID	LPIPS	SIFID	LPIPS	SIFID	LPIPS
SinGAN	0.08	0.68	0.12	0.26	0.29	0.26	0.72	0.09
GPNN	0.32	0.39	0.34	0.18	0.61	0.18	0.51	0.07
ConSinGAN	0.10	0.55	0.10	0.25	0.47	0.25	0.31	0.27
SinDDM	0.36	0.30	0.39	0.11	0.57	0.24	0.72	0.07
SinDiffusion	0.42	0.33	0.37	0.13	0.63	0.16	0.71	0.12
CharGAN	**0.04**	**0.71**	**0.07**	**0.30**	**0.11**	**0.32**	**0.09**	**0.35**

The best values are highlighted in bold. For SIFID, lower is better, and for LPIPS, higher is better

Table 2 Comparison of PSNR and SSIM metrics for images generated by various single image generative models

Method	**Stone**		**Palm leaf**		**AHCD**		**EMNIST**	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
SinGAN	30.32	0.42	33.29	0.31	33.87	0.56	32.01	0.51
GPNN	36.89	0.70	44.70	**0.71**	39.49	0.70	32.70	0.63
ConSinGAN	33.89	0.60	34.62	0.53	36.97	0.44	32.22	0.40
SinDDM	30.45	0.43	32.83	0.26	33.09	0.53	31.30	0.34
SinDiffusion	30.45	0.41	33.29	0.30	33.58	0.40	31.66	0.47
CharGAN	**37.01**	**0.73**	**46.31**	0.67	**42.07**	**0.89**	**34.02**	**0.76**

The highest values are highlighted in bold

models in terms of generating quality and diversity. We use SIFID and LPIPS to assess image quality and diversity, respectively. SIFID [6, 32] quantifies the feature-space distance between the generated and original images. LPIPS [33] compared perceptual differences using a pre-trained AlexNet [34], which is color-insensitive yet spatially sensitive. Although some approaches yield excellent LPIPS scores, they perform poorly for SIFID and show poor visualization results. Compared to other methods, the proposed method achieves low SIFID and high LPIPS scores simultaneously, indicating that CharGAN can generate diverse samples with natural structures. In Section "Qualitative Results", we discuss the performance of various models in the context of character datasets, particularly highlighting the effectiveness of pyramidal structure models, such as SinGAN, GPNN, and our proposed model. However, upon further analysis, it became apparent that the GPNN performance may not consistently outshine other methods across all metrics or datasets.

A comparative analysis of generative models based on the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) in Table 2 reveals the exceptional performance of CharGAN across multiple datasets. CharGAN consistently achieves the highest PSNR values, indicating superior noise reduction capabilities, with notable scores such as 37.01 for the Stone dataset, 46.31 for the Palm leaf dataset, 42.07 for the AHCD dataset, and 34.02 for the EMNIST dataset. Additionally, CharGAN excelled in structural preservation, achieving the highest SSIM values for most datasets, including 0.73 for Stone, 0.67 for Palm Leaf, 0.89 for AHCD, and 0.76 for EMNIST. While the GPNN shows competitive SSIM performance on the palm leaf dataset, CharGAN's overall dominance in both metrics underscores its robustness and effectiveness. These results highlight the capability of CharGAN to produce high-quality, structurally similar images across diverse datasets, making it a promising choice for various applications.

Comparative analysis of parameter complexity in single image generative models

Table 3 compares the parameters obtained using the proposed method with other single-image generative models. The analysis of parameter counts among the various generative models highlights the distinct positioning of the proposed CharGAN model. The proposed CharGAN model, with around 9.2 × 10^5 parameters, offers a balanced complexity between simpler models like ConSinGAN ($\sim 6.1 \times 10^5$) and more complex ones such as SinDDM ($\sim 1.1 \times 10^6$) and SinDiffusion ($\sim 1.3 \times 10^6$). However, CharGAN aims to achieve a middle ground, providing robust performance without the excessive complexity observed in the highest-parameter models. This balance makes CharGAN a promising candidate for applications requiring efficient yet powerful generative capabilities. GPNN is not included in this comparison because it relies on a non-parametric approach to image generation.

Fig. 7 is a bar chart comparing the number of parameters for the different models. The chart uses a logarithmic scale on the y-axis to illustrate the differences better. CharGAN, the proposed model, is shown in red, which balances the complexity between simpler and more complex models.

Ablation study

We used the SinGAN as a baseline and performed ablation experiments to examine the impact of each component on the generation process. We investigated

Table 3 Number of parameters for different models

Model	**Parameters**
SinGAN	$\sim 9.8 \times 10^5$
ConSinGAN	$\sim 6.1 \times 10^5$
SinDDM	$\sim 1.1 \times 10^6$
SinDiffusion	$\sim 1.3 \times 10^6$
CharGAN	$\sim 9.2 \times 10^5$

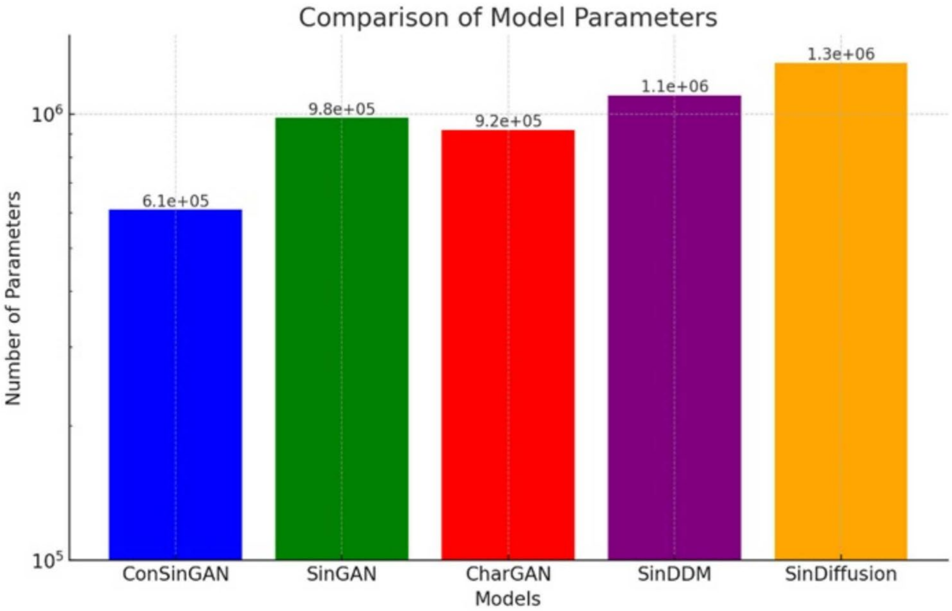

Fig. 7 Comparison of model parameters for the different single image models

Table 4 Quantitative results of Ablation study

Method	Stone		Palm leaf		AHCD		EMNIST	
	SIFID	**LPIPS**	**SIFID**	**LPIPS**	**SIFID**	**LPIPS**	**SIFID**	**LPIPS**
SinGAN	0.08	0.68	0.12	0.26	0.29	0.26	0.72	0.09
SinGAN+A	0.06	0.61	0.09	0.23	0.18	0.21	0.24	0.17
SinGAN+L	0.12	**0.72**	0.10	0.28	0.22	0.30	0.59	0.28
SinGAN+A+L	**0.04**	0.71	**0.07**	**0.30**	**0.11**	**0.32**	**0.09**	**0.35**

The best values are highlighted in bold. For SIFID, lower is better, and for LPIPS, higher is better

the effects of two changes to the basic SinGAN architecture: the augmentation inducer and the complicated prior input to the generator. We quantified the influence of each component using SIFID and LPIPS, individually and then in combination, as shown in Table 4.

In the Table, 'SinGAN+A' represents the performance after applying the Augmentation inducer to the basic SinGAN model for four different datasets, and 'SinGAN+L' represents the performance after supplying the generator random sampling from the dynamic Gaussian mixture model. The two enhancement strategies were then combined (denoted as 'SinGAN+A+L') to test their performance when combined, which is the entire version of our proposed model.

Conclusion and future work

We have introduced a guided unconditional single-image generative model named CharGAN for paleographic and other handwritten character images. CharGAN used dynamic sampling from a Gaussian mixture model as input to the generator to change the minute features of the generated images and an Augmentation inducer to change the high-level characteristics. Supporting this, CharGAN generates images appropriate for highly similar characters while preserving the global structure and creating realistic and diverse samples. In addition, we experimented on four different handwritten character datasets, which show that our approach outperforms other single-image generative models in quantitative and qualitative terms.While our study focused on four specific datasets, it's important to emphasize that the proposed method holds promise for application across a wide range of scripts, including Chinese, Korean, Japanese, ancient Egyptian, and others with similar characteristics. By acknowledging the potential applicability of our method to a diverse array of scripts, we aim to encourage future research endeavors that explore its efficacy in various linguistic and cultural contexts. Expanding the scope beyond paleographic studies, the character synthesis based on the unconditional frame model presented in our article holds significant potential for various applications. Beyond the realm of paleography, this method can find utility in fields such as digital art, font design, and educational tools for language learning.

Abbreviations

GAN	Generative adversarial network
DCGAN	Deep convolutional GAN
WGAN	Wasserstein GAN
DDPM	Denoising diffusion probabilistic models
SIFID	Single image frechet inception distance
LPIPS	Learned perceptual image patch similarity
GPNN	Generative patch nearest neighbors
GPDM	Generative patch distribution matching
SinGAN	Single-image-GAN
ConSinGAN	Concurrent-single-image-GAN
ExSinGAN	Explainable single image generative model
SIV-GAN	Single image and video GAN
SinDiffusion	Single-image diffusion model
SinDDM	Single image denoising diffusion model
FS	Conv fractional-strided convolution
BN	Batch normalization
LReLU	Leaky ReLU
ASI	Archaeological survey of India
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity index measure

Acknowledgements

None.

Author contributions

AA conceived the concept, collected data, made calculations, and wrote the text with assistance from UM. UM is responsible for designing and verifying analytical procedures, as well as supervising the outcomes of work. Both authors reviewed and approved the final paper.

Funding

None.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 8 February 2024 Accepted: 16 July 2024
Published online: 26 July 2024

References

1. Lombardi F, Marinai S. Deep learning for historical document analysis and recognition-a survey. J Imaging. 2020;6(10):110.
2. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020;63(11):139–44.
3. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv. 2015. https://doi.org/10.4855/arXiv.1511.06434.
4. Arjovsky MC. S. & Bottou, L. Wasserstein GAN. Proceedings of ICML 2017. 2017;.
5. Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. arXiv preprint. 2018. https://doi.org/10.4855/arXiv.1809.11096.
6. Shaham TR, Dekel T, Michaeli T. Singan: Learning a generative model from a single natural image. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019. p. 4570–4580.
7. Hinz T, Fisher M, Wang O, Wermter S. Improved techniques for training single-image gans. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2021. p. 1300–1309.
8. Zhang Z, Han C, Guo T. ExSinGAN: learning an explainable generative model from a single image. arXiv preprint. 2021. https://doi.org/10.48550/arXiv.2105.07350.
9. Sushko V, Gall J, Khoreva A. Learning to generate novel scene compositions from single images and videos. arXiv preprint. 2021. https://doi.org/10.48550/arXiv.2105.05847.
10. Granot N, Feinstein B, Shocher A, Bagon S, Irani M. Drop the gan: In defense of patches nearest neighbors as single image generative models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 13460–13469.
11. Elnekave A, Weiss Y. Generating natural images with direct patch distributions matching. Berlin: Springer; 2022. p. 544–60.
12. Nikolaidou K, Seuret M, Mokayed H, Liwicki M. A survey of historical document image datasets. Int J Document Anal Recognit (IJDAR). 2022;25(4):305–38.
13. Shocher A, Bagon S, Isola P, Irani M. Ingan: Capturing and retargeting the "dna" of a natural image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019. p. 4492–4501.
14. Nikankin Y, Haim N, Irani M. Sinfusion: training diffusion models on a single image or video. arXiv preprint. 2022. https://doi.org/10.48550/arXiv.2211.11743.
15. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Adv Neural Inform Proc Syst. 2020;33:6840–51.
16. Wang W, Bao J, Zhou W, Chen D, Chen D, Yuan L, et al. Sindiffusion: learning a diffusion model from a single natural image. arXiv preprint. 2022. https://doi.org/10.4855/arXiv.2211.12445.
17. Kulikov V, Yadin S, Kleiner M, Michaeli T. Sinddm: A single image denoising diffusion model. In: International Conference on Machine Learning. PMLR; 2023. p. 17920–17930.
18. Kingma DP, Salimans T, Welling M. Variational dropout and the local reparameterization trick. Adv. Neural Inf. Process. Syst., vol. 2015-Janua, no. Mcmc, pp. 2575–2583.
19. Sindagi VA, Patel VM. Generating high-quality crowd density maps using contextual pyramid cnns. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 1861–1870.
20. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pmlr; 2015. p. 448–456.
21. Thomson W, Jabbari S, Taylor A, Arlt W, Smith D. Simultaneous parameter estimation and variable selection via the logit-normal continuous analogue of the spike-and-slab prior. J Royal Soc Interface. 2019;16(150):20180572.
22. Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK. An introduction to variational methods for graphical models. Mach Learn. 1999;37:183–233.
23. Jordan MI. Learning in graphical models. Cambridge: MIT press; 1999.
24. Rezende D, Mohamed S. Variational inference with normalizing flows. In: International conference on machine learning. PMLR; 2015. p. 1530–1538.
25. Kingma DP, Salimans T, Jozefowicz R, Chen X, Sutskever I, Welling M. Improved variational inference with inverse autoregressive flow. Advances in neural information processing systems. 2016;29.
26. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint. 2014. https://doi.org/10.4855/arXiv.1412.6980.
27. Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14. Springer; 2016. p. 702–716.

28. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1125–1134.
29. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of wasserstein gans. Advances in neural information processing systems. 2017;30.
30. AlJarrah MN, Mo'ath MZ, Duwairi R. Arabic handwritten characters recognition using convolutional neural network. In: 2021 12th International Conference on Information and Communication Systems (ICICS). IEEE; 2021. p. 182–188.
31. Cohen G, Afshar S, Tapson J, Van Schaik A, EMNIST: Extending MNIST to handwritten letters. In,. international joint conference on neural networks (IJCNN). IEEE. 2017;2017:2921–6.
32. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium, Adv. Neural Inf. Process. Syst., vol. 2017-Decem, no. Nips, pp. 6627–6638. https://doi.org/10.18034/ajase.v8i1.9.
33. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 586–595.
34. Gonzalez TF. Handbook of approximation algorithms and metaheuristics. Boca Raton: CRC Press; 2007.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.