Content-Aware Urdu Handwriting Generation

Zeeshan Memon[1], Adnan Ul-Hasan[2], and Faisal Shafait[1,2]

[1] School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad, Pakistan
[2] Deep Learning Laboratory, National Center of Artificial Intelligence (NCAI), Islamabad, Pakistan
{zeeshan.bese20seecs,adnan.ulhassan,faisal.shafait}@seecs.edu.pk

Abstract. The performance of handwriting recognition systems has undergone significant improvement in recent years. However, the accuracy of these systems for multiple cursive scripts, including Arabic and Urdu, is still limited due to the lack of labeled training data. Handwriting generators are a potential solution to this problem. Previous research on Urdu handwriting generation has primarily focused on generating realistic ligatures using Generative Adversarial Networks (GANs) with common adversarial loss but has not addressed the issue of maintaining content and generated image entanglement. This paper aims to address this gap by proposing a content-controlled training approach for Urdu Handwriting Generation with pre-trained recognizer loss. Our generation model is trained on a diverse set of printed ligatures and then fine-tuned with transfer learning on handwritten images. In this paper, a new metric for evaluating the performance of handwriting generation systems is also suggested, which is specifically tailored to the context of handwriting generation tasks. To our knowledge, this is the first Urdu handwriting generation system that is capable of generating content-controlled images.

Keywords: Handwriting Generation · Recognition Loss · Generated Adversarial Network

1 INTRODUCTION

Handwriting is a fundamental aspect of human communication and has played an important role in the documentation of human history. From ancient civilizations to the present day, handwriting has been used to record personal and public events, preserving them for future generations. The use of digital devices has made it easier and more efficient to produce and share written materials. Moreover, more people now rely on typing rather than handwriting to create written documents [10]. Handwriting is still used today for record-keeping in certain fields such as medicine and therapy [12], where it is important to have legible, accurate records of patient information. Additionally, handwriting is also significantly used in education as teachers and students often take notes by hand and it is seen as a personal expression.

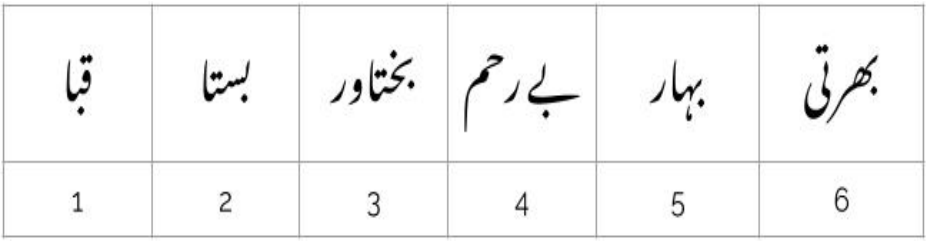

Fig. 1. The figure illustrates the contextual variations of the Urdu character '*bay*' based on its position (initial, medial, or final) and its combination with other characters [2].

Optical Character Recognition systems (OCRs) have achieved great performance against printed text but still lacks behind in handwritten text due to limited data and diverse writing styles, specifically for Arabic scripts including Urdu.

Urdu, Farsi, Sindhi, Pashto, and Punjabi are all written in scripts that are derived from the Arabic script, which has unique characteristics due to its cursive nature [1]. The shape of characters within the script is dependent on their position within a given word as shown in Fig. 1. Additionally, the use of diacritics and dots serves to indicate grammar and pronunciation [1]. There also exists diverse variations when it comes to inter-word and intra-word spacing within the script with overlapping characters, which adds more to its complexity as highlighted in Fig. 2.

Furthermore, Arabic script is cursive in nature, where a single word consists of one or more ligatures. A ligature is a combination of two or more characters that are merged into a single more complex shape [11]. These factors, in conjunction with individual variations in handwriting style, contribute to the complexity of processing and analyzing text written in the Arabic script.

Given the challenges posed by complex and limited data, it has become increasingly evident that the current recognition systems are not meeting the desired standards of performance. Recently, there has been a growing interest in the field of handwriting generation as a potential solution for the data limitation challenge. Variational Autoencoders (VAEs) [19] and Generative Adversarial Networks (GANs) have emerged as popular research areas in this field. However, a major challenge in these studies, particularly for Arabic scripts, is the absence of controlled text generation [5]. Controlled text generation is a critical aspect of handwriting generation for OCR systems, as it is necessary for training OCR models and obtaining annotated data that accurately reflects the needs of the OCR systems [18].

From the current literature, two significant research gaps have been identified. Firstly, the lack of content-controlled generation for complex script handwriting is a major challenge [10]. This means that there is difficulty in generating images that accurately represent the intended content when dealing with scripts that have a high level of complexity. Secondly, even when content control is attempted, the concurrent training of the recognition system with the GANs presents an-

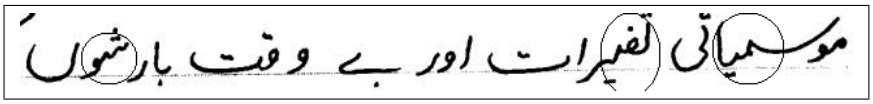

Fig. 2. Illustrating the Complexity of Urdu Handwriting: A Sample of Overlapping Characters in Urdu Script. The image shows the cursive and overlapping nature of the Urdu script.

other challenge. This is because the recognition system can validate substandard generated images as readable, which does not ensure that the generator will converge towards producing realistic and readable handwriting images [20]. This can result in the generation of images that do not accurately represent the intended content.

In this paper, an alternative approach is proposed that presents a novel solution to the challenges of content-controlled and readable generation of handwriting images. By incorporating a pre-trained recognizer network into a generative end-to-end architecture, the generational power of GANs is leveraged while addressing the issue of model failure in producing readable text images. For any previously unseen unicode Urdu string given as input, our approach generates an image of the corresponding Urdu text in a handwriting style.

This paper is further divided into different sections. Section 2 summarizes the relevant work done previously. Section 3 provides a detailed methodology of the proposed approach. Section 4 presents experimental configurations and Section 5 discusses the results and compares them with other similar works. Section 6 finally concludes the paper with a summary and future directions.

2 PREVIOUS WORK

Handwritten text generation techniques utilizing deep learning can be broadly classified into two categories: online and offline generation techniques. Online techniques typically utilize temporal data obtained from the sequential recording of real handwriting samples (in vector form) via the use of a digital stylus [4]. Alternatively, recent generative offline handwritten text generation techniques [5] focus on the direct generation of text through training on offline handwriting images.

Graves et al. [3] present the very first approach utilizing a Recurrent Neural Network (RNN) with Long-Short-Term Memory (LSTM) cells for predicting future stroke points based on previous pen positions and input text. Further, Ji et al. [6] extended the method presented in [3] by incorporating a GAN framework with a discriminator. The introduction of a disentanglement mechanism in DeepWriting [4] allows for greater control over the generation of style without affecting the content. Haines et al. [7] proposed a method for author-specific handwriting generation, which requires a significant amount of character-level annotation for each new sample.

Recent advancements in handwriting generation have aimed to improve control over both content and style. One such approach is presented by Alonso et

al. [5], which utilizes a GAN architecture composed of a discriminator and generator, as well as two additional networks: a bidirectional LSTM and a CNN with LSTM layers at the end. This approach specifically focuses on the generation of fixed-length and width handwritten strings in French and Arabic. The generated images were incorporated into an existing dataset, resulting in improved accuracy of recognition systems.

In another study, Fogel et al. [8] presents a novel method for generating images from text, referred as ScrabbleGAN, which incorporates a correlation between the width of the generated image and the length of the input text. The results of ScrabbleGAN demonstrate its proficiency in generating high-quality images that are semantically consistent with the input text. Farooqui et al. [9] presents an approach for improving handwriting recognition of the Urdu language by generating additional data samples using different GAN variants. Seven different GAN architectures were implemented for the generation of handwritten Urdu ligatures, with each GAN trained to produce a specific class of ligatures or at most 10 classes for class-conditioned variants. The goal is to increase the amount of training data to improve the accuracy of word-spotting tasks. Sharif et al. [10] present a GAN-based approach for Urdu Handwriting Generation, which produces realistic Urdu ligatures. Three different GANs variants were evaluated, with WGANs showcasing the best performance. It basically utilizes the receptive power of a deep convolutional generator to generate complex overlapping ligatures, but it does not ensure content-controlled generation.

Similar to ScrabbleGAN [5], we also investigate the problem of content-controlled generation and propose an approach, utilizing a pre-trained recognizer network with frozen weights during training instead of training along with GANs, to ensure a stable mapping between input character embeddings and generated images. This approach is intended to overcome the limitations present in current handwriting generation techniques.

3 PROPOSED METHODOLOGY

In this study, we propose to use a GAN-based approach as shown in Fig. 3, where in addition to the discriminator, the generated image is also evaluated by a recognizer network. The purpose of the discriminator is to promote the realistic appearance of handwriting styles, while the recognizer network serves to ensure the generated image is readable and accurately represents the input text. Each component is discussed in the following subsections.

3.1 Fully Convolutional Generator

The fundamental principle guiding our proposed model is the realization that handwriting is a localized process, meaning that each letter is influenced only by the letters preceding and succeeding it. This is also supported by Graves et al. [3], who employed recurrent neural networks for the task at hand.

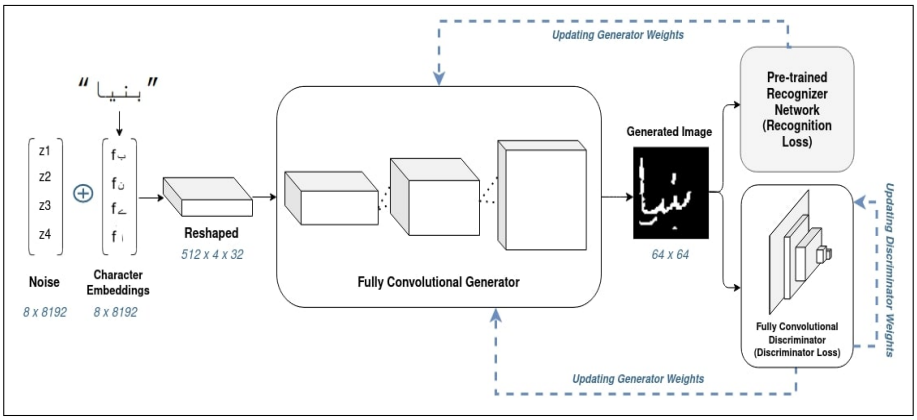

Fig. 3. Proposed approach for Ligature Image Generation. Embeddings for each character in a given ligature are combined with noise before being fed into the network. The image is generated through upsampling and convolutional layers and then passed through a convolutional network and recognizer network, resulting in discriminator loss and recognition loss, which guide the learning of the overall architecture (as shown by the dashed blue line).

By analyzing Urdu ligature formation, we posit that the shape of characters in Urdu is heavily influenced by the surrounding characters within the ligature. This suggests that this characteristic can be effectively learned through the use of convolutional neural networks. The proposed generator utilizes character embeddings and maps them onto the generated image through a series of convolutional layers.

The generator can be conceptualized as one that generates individual character-wise patches, rather than generating complete words in their entirety. A combination of convolutional layers and upsampling layers in each layer of the generator is employed which increases the overlap between neighboring characters, thereby expanding the receptive field. This facilitates interactions among neighboring characters, resulting in a smoother transition within ligatures. For every character in a given ligature, a character embedding is combined with a noise vector in order to account for natural variations in handwriting. The resulting embeddings are then passed through a fully convolutional generator, where the region generated by each character filter is of the same dimension and the receptive fields of adjacent filters overlap to generate the ligature image.

3.2 Fully Convolutional Discriminator

In the traditional GAN architecture, the role of the discriminator is to accurately distinguish between original data samples and those generated by the generator. In the proposed model, a discriminator is also utilized to score images as either real or fake. The discriminator is a fully convolutional neural network, similar

to the generator, but with an architecture opposite of the generator. Both actual handwritten samples and generated samples are provided as input to the discriminator, which then evaluates these images and produces an output. This output is subsequently used in the loss function to update the weights of both the generator and discriminator.

3.3 Recognizer Network

The recognizer network evaluates generated images on the basis of readability by comparing the recognized text from the recognizer network with the input label provided to the generator. For the recognizer network, we have used the state-of-the-art recognition system for Urdu Handwriting proposed by Riaz et al. [12], which combines the capabilities of a convolutional neural network (CNN) and a transformer (Conv-Transformer). The CNN component extracts visual features from the input image, which are then passed to a full transformer consisting of three encoder-decoder layers. The model employs a cross-entropy loss function to measure the difference between the predicted text and the text labels.

For our task, we trained the Conv-Transformer architecture as suggested by Riaz et al. [12] on NUST-UHWR Dataset [18] with the objective of achieving generalizability of recognizer network.

3.4 Optimization Functions

Three distinct learning objectives are discussed in this work. Specifically, discriminator loss, generator loss, and recognition loss are utilized. The utilization of these three objectives aims to enforce the content-controlled generation of images, thus enhancing their overall quality.

Discriminator Loss We employ a discriminator model to estimate the probability of whether a given sample is from the training data (X) or from the artificially generated distribution. The optimization problem is formulated as a min-max problem, where the generative network (G) and the discriminator (D) are trained in competition with each other. Formally, it can be defined as

$$\min_{L_D} L_D = D((G(Z, L)), 0) - D(X, 0)) \tag{1}$$

where Z and L represent noise vector and text label. The G represents the generator network, which generates text images given Z and L.

Generator Loss It is typically a function that measures how realistic generated images are. It can be formally defined as

$$\min_{L_G} L_G = -D((G(Z, L)), 0) \tag{2}$$

Recognition Loss A pre-trained state-of-the-art Urdu recognition is utilized as a handwritten text recognizer network (R) that guides the generation of synthetic

Fig. 4. Text, its rendered and augmented Versions, where **(a)** displays the input text, **(b)** presents the rendered version using Pango with the 'Pak Nastaleeq'font and **(c)** showcases additional augmentations that mimic real handwriting variations.

word images with specific textual content. As the recognizer network is frozen during training, this loss optimizes the weights of the generator network only. This is a fundamental minimization problem, which can be defined as

$$\min_{L_R} L_R = R((G(Z, L)), L) \tag{3}$$

The overall architecture is trained using a combination of three proposed loss functions keeping the weights of the recognizer network freeze. The three losses represented by equations (1), (2), and (3) are combined arithmetically to yield the overall loss.

$$\min_{L} L = L_D + L_G + L_R \tag{4}$$

The weights of the generator and discriminator are updated in an alternating fashion to ensure the stability of the overall learning process.

4 EXPERIMENT CONFIGURATION

Several experiments were conducted to evaluate the effectiveness of the proposed model for generating Urdu ligatures and to compare its results with those of existing baselines and state-of-the-art approaches. The specifications of the datasets used implementation details, and hyperparameter settings are thoroughly discussed below.

4.1 Datasets Used

In order to assess the performance of the proposed model, both rendered and real handwriting datasets were utilized. A brief overview of the database is provided below:

UCOM Database: In order to evaluate the proposed network, the UCOM database [13] is utilized. The dataset consists of 48 distinct lines of Urdu text, authored by 100 different individuals. The Urdu language comprises 36 unique alphabets, with standalone Urdu alphabets also being considered as ligatures of a single character. Following the methodology outlined in [9], the 317 unique ligatures are extracted from images of Urdu sentences through binarization, segmentation, and resizing to obtain ligatures of a fixed dimension. By utilizing data augmentation 32,000 samples were obtained.

Center of Language Engineering (CLE) database: The CLE database [23], developed by the Center for Research in Urdu Language Processing (CRULP) [24], primarily consists of 18,000 frequent Urdu ligatures in Unicode format. These ligatures are organized based on the number of characters, ranging from 2 to 8 characters. We used ligatures consisting of up to 4 characters, yielding a total of 10,012 unique ligatures. These ligatures are rendered using Pango [14] as shown in Fig. 4.

4.2 Pre-processing of Dataset

We argue that better formation of ligatures can be learned through a large corpus of ligatures, and this knowledge can be transferred and refined to a specific handwriting dataset. In order to achieve this, augmentations that aim to make the rendered images as similar as possible to real handwriting are utilized. These augmentations include erosion, dilation, rotation, and shear transformation as shown in Fig. 4. A dataset of 30,036 images is generated by using these augmentations on 10,012 unique ligatures from the CLE database.

4.3 Implementation Details and Hyper Parameters

The architecture of the network is configured to generate fixed-size images of 64×64 pixels. The input ligature is padded up to a sequence length of eight characters and then passed through the embedding layer of *Generator (G)* to generate embeddings of shape 8×8192 for each sample. As illustrated in Fig. 3, for the generation of an n-character ligature, 'n'number of character embeddings are generated according to the characters. These embeddings are reshaped into $512 \times 4 \times 32$ and subsequently passed through convolutional layers, followed by an upsampling layer. Leaky ReLU (LReLU) and batch normalization [15] is applied between these layers, and a sigmoid activation function is utilized to produce the final output of size 64×64. Table 1 shows the detailed architecture of the generator with corresponding output shapes.

The *Discriminator (D)* network is essentially the inverse of the generator network, with the exception of the absence of the spatial embeddings layer. An image of 64×64 pixels is provided as input to the discriminator, which is then processed through a series of layers, including the convolutional layer, Leaky ReLU (LReLU) layer, batch normalization, and max pool layer. The final layer

is a linear layer that outputs a single value, representing the score or probability of the image being real or fake. The detailed architecture of the discriminator, encompassing dimensions and activation functions at each layer, is presented in Table 1. A batch size of 32 was employed, where the input labels' sequence length was padded up to the maximum sequence length of 8. The Adam optimizer with a learning rate of $2e^{-4}$ was utilized for the training of our architecture. For every generator update, the discriminator is updated 7 times, and recognition loss is optimized on every 5th training step of the epoch.

Hyperparameter tuning The stability of GAN training is a well-known challenge in the field of generative modeling. The stability of GANs is influenced by several factors, including the learning rate and the number of times the discriminator is trained compared to the generator, as stated in the seminal work by [16]. In practice, there is no standard set of hyperparameters that works for all models and datasets.

In our study, we conducted an extensive exploration of the hyperparameters to stabilize the training of GANs. The learning rate was varied from $1e^{-4}$ to $1e^{-5}$ with intervals, while the number of times the discriminator was trained relative to the generator, represented by the parameter 'k', was varied from 2 to 10. Our results showed that a learning rate of $2e^{-5}$ and a value of k=7 were the optimal hyperparameters for our proposed approach.

4.4 Experiments Performed

We have executed three distinct variations of experiments, considering the dataset or approach employed. Our results have been benchmarked against the state-of-the-art in Urdu Handwriting Generation [10] and are thoroughly discussed in subsequent sections.

Performance on CLE database The proposed model was trained on the CLE database from scratch, utilizing 300 epochs with hyperparameters in accordance with the specifications described in Section 4.3. However, the discriminator was trained seven times that of the generator, as advised in [16]. Out of 30,036 samples, 28,512 were designated for training and the rest for testing. The effectiveness of the proposed model was evaluated using the Fréchet Inception Distance (FID), Geometric Score (GS), and Recognition Accuracy, as outlined in Table 2. The Recognition Accuracy was determined through the Character Accuracy Rate, which reflects the number of characters that can be recognized from the generated images through the use of a state-of-the-art Urdu handwriting recognition system.

Performance on UCOM database In a similar fashion to the training on the CLE database, the proposed model underwent 300 epochs of training with hyperparameters consistent with those specified for the CLE database, except for

Table 1. Configurations of custom Generator and Discriminator blocks. Each convolution layer has ReLU activation except the last one which has Sigmoid Activation.

Generator layer	**Output shape**	**Discriminator layer**	**Output shape**
Embedding Layer + Noise	8×8192	Input Image Vector	$1 \times 64 \times 64$
Embedding Layer + Noise	$512 \times 4 \times 32$	Convolution	$32 \times 64 \times 64$
(Reshaped)		Convolution	$32 \times 64 \times 64$
Convolution	$256 \times 8 \times 32$	Convolution	$32 \times 64 \times 64$
Batch Normalization	$256 \times 8 \times 32$	Convolution	$64 \times 32 \times 32$
Convolution	$128 \times 16 \times 32$	Convolution	$128 \times 16 \times 16$
Batch Normalization	$128 \times 16 \times 32$	Convolution	$128 \times 16 \times 16$
Convolution	$128 \times 32 \times 32$	Convolution	$256 \times 16 \times 8$
Batch Normalization	$128 \times 32 \times 32$	Batch Normalization	$256 \times 16 \times 8$
Convolution	$64 \times 64 \times 64$	Convolution	$256 \times 16 \times 4$
Convolution	$64 \times 64 \times 64$	Batch Normalization	$256 \times 16 \times 4$
Convolution	$32 \times 64 \times 64$	Convolution	$256 \times 16 \times 4$
Convolution	$32 \times 64 \times 64$	Batch Normalization	$256 \times 16 \times 4$
Convolution	$16 \times 64 \times 64$	Convolution	$256 \times 16 \times 2$
Convolution	$1 \times 64 \times 64$	Linear Layer	1×1

the discriminator which was trained five times than that of the generator. The performance of the model was assessed using the FID score, Geometric Score, and Recognition Accuracy, as shown in Table 3.

Performance of Model trained on CLE and UCOM database The processed UCOM database consists of only 317 unique ligature formations as highlighted in Section 4.1, which makes it insufficient as an Urdu Handwriting Generator due to its limited data. To improve the dataset, 10,000 ligatures from the CLE database have been rendered and augmented to incorporate real handwriting variations. The model was trained on the CLE data and then transferred and fine-tuned on the UCOM database for an additional 50 epochs with a learning rate of $2e^{-6}$, drawing inspiration from the transfer learning approach used in GANs [17]. The performance was evaluated using the FID score, Geometric Score, and Recognition Accuracy, and the results are presented in Table 2.

Impact of Generated Data on Urdu OCR Performance The objective of handwriting generation is to increase annotated data to improve handwriting recognition accuracy. To evaluate the improvement, an experiment was performed in which the OCR model was trained with both the generated data and the UCOM database. The performance of the proposed model was compared to Sharif et al. [10] using the Character Error Rate (CER) as the evaluation metric, shown in Table 4.

5 RESULTS AND DISCUSSION

The performance of our proposed method was evaluated using three quantitative metrics: Fréchet Inception Distance (FID) [22], Geometric Score (GS) [21], and Recognition Accuracy. FID was utilized to measure the similarity between the feature representations of the generated and real images. This was achieved by fitting two Gaussians on the feature representations obtained from an Inception Network and calculating the Fréchet distance between them. GS, on the other hand, compares the geometrical properties of the fundamental real and fake data manifolds and provides a means to quantify mode collapse.

In this study, a new evaluation metric, Recognition Accuracy, has been discussed as a more effective means of evaluating text image generation tasks. Unlike FID and Geometric Score, which evaluate the generated text images based on latent features, Recognition Accuracy evaluates the readability of the images through state-of-the-art generalized OCR systems. Although there may be limitations to this approach due to the limitations of OCR itself, it provides a standardized means of determining the quality of text images. Additionally, the FID score has its own limitations, as the Inception network it relies on is primarily trained on facial data, which may not accurately represent the target distribution for another target. For this reason, the use of Recognition Accuracy along with the FID score may provide a more comprehensive evaluation of the quality of content-controlled handwriting generation tasks.

In every experiment performed, the FID score was calculated for the entire dataset, comparing it with an equivalent number of generated samples, approximately 30,000 in total. The Geometric Score was determined through the analysis of 5,000 real and 5,000 generated samples using default parameter settings. Furthermore, the Recognition Accuracy was calculated for the complete dataset using a pre-trained recognizer network.

5.1 Results on CLE Database

The results of our study on the CLE database show an FID score of 69.01 and a Geometric Score of $7e^{-4}$, with a recognition accuracy of 77% as highlighted in Table 2. The recognition accuracy of 77% indicates that 77% of characters are readable in our generated images with the help of generalized OCR. Given that the CLE database is comprised of 10,000 unique ligatures and trained for 300 epochs, the results demonstrate the good performance of the proposed model.

Fig. 5. Label, it's ground truth (printed text with Pango) from **CLE database** and generated sample through the proposed model, where **(a)** displays the input text, **(b)** presents ground truth and **(c)** showcases generated samples through the proposed model.

Table 2. Comparison of FID score, Geometric Score and Recognition Accuracy for proposed approach on **CLE database** and fine-tuned on **UCOM database**.

Training Data	**FID Score**	**Geometric Score**	**Recognition Accuracy(%)**
CLE Database	69.01	$7.87e^{-4}$	77
CLE Database + Fine-tuned on UCOM database	38.03	$8.81e^{-4}$	72.6

Furthermore, a qualitative assessment as shown in Fig. 5 confirms the validity of the results. The utilization of an augmented training dataset ensures that the generated images accurately depict natural handwriting, and the formation of ligatures in comparison with ground truth confirms the correctness of the model.

5.2 Results on UCOM database

Results of the proposed approach on the UCOM database as explained in 4.4 demonstrate an FID score of 23.24 and a Geometric Score of $5.95e^{-}4$, with a recognition accuracy of 69.7% as illustrated in Table 3. The quality and accuracy of the images produced can also be confirmed by the visual representation in Fig. 6. As demonstrated, the majority of the generated samples contain recognizable characters, with the exception of one most right sample which is a five-letter ligature, and one out of five characters not being recognizable based on human evaluation. These results align with the quantitative recognition accuracy mentioned in Table 3, which is a mislabeling rate of 4-5% in the dataset and a limitation of the recognition model's accuracy.

Along with the quantitative comparison presented in Table 3, a qualitative analysis was also conducted, as shown in Fig. 7. The results demonstrate that the

Fig. 6. Label, its ground truth from **UCOM database** and generated sample through proposed model, where **(a)** displays the input text, **(b)** presents ground truth and **(c)** showcases generated samples through proposed model.

Fig. 7. Qualitative comparison of Proposed Model and different GAN variants from Sharif et al. [10] on UCOM database. **(a)** shows ground truth, **(b) - (d)** represents generated samples from Deep Convolutional GANs(DCGANs), Wasserstein GANs (WGANs), and Wasserstein GANs with gradient penalty (WGANs-GP), **(e)** showcase generated samples through the proposed model.

proposed model generates samples that are comparable in quality to those generated by other GAN variants from Sharif et al. [10]. It is worth noting that the proposed model generates content-controlled samples, while the samples from the previous works were the best-generated samples of the same class/label sep-

Fig. 8. Qualitative comparison of generated samples and ground truth when transfer learning is employed from **CLE database** to **UCOM database**, where **(a)** displays the input text, **(b)** presents ground truth and **(c)** showcases generated samples through the proposed model.

Table 3. Comparison of FID score, GS, and Recognition Accuracy for different GAN variants from previous works and Proposed Approach.

Model	**FID Score**	**Geometric Score**	**Recognition Accuracy(%)**
DCGANs	21.45	$7.82e^{-4}$	35.1
WGANs	17.97	$7.46e^{-4}$	37.6
WGANs-GP	**15.74**	$7.14e^{-4}$	39.2
Proposed Model	23.24	$5.95e^{-4}$	**69.7**

arated manually as they are not content-controlled generation. This superiority of the proposed model can be verified by examining Table 3, which shows no significant difference in FID and geometric score but a marked improvement in recognition accuracy, reflecting the difference between controlled and uncontrolled generation. The results highlight the ability of the proposed model to generate annotated samples of equivalent quality to those generated by uncontrolled methods while still maintaining a close relationship to the input.

5.3 Result of Transfer Learning on UCOM database

As detailed in Section 4.4, an attempt was made to apply transfer learning from large rendered and highly augmented data (CLE database) to real handwriting data, with the aim of improving ligature formation and increasing the generalization of the generator. The results are presented in Fig. 8, which shows that while the difference between generated samples and ground truth can still be distinguished, the generator is making progress toward replicating the smoothed strokes of real handwriting images. Table 2 reports the evaluation metrics, including the FID score of 38.03, the Geometric Score of $8.81e^{-4}$, and recognition accuracy of 72.6%, demonstrating the efficacy of this approach. These results

Table 4. Comparison of OCR performance in terms of CER(Character Error Rate), when trained with Synthetic data generated with WGANs-GP [10] and proposed model.

Training Data	**Model**	**CER(%)**
UCOM	-	7.12
UCOM + Generated Data(10k Samples)	Sharif et al. [10]	6.77
UCOM + Generated Data(10k Samples)	**Proposed Model**	**6.15**

are also in line with those obtained from training a model from scratch using only the UCOM database, as shown in Table 3. This highlights the potential of transfer learning in handwriting generation tasks when the training data is limited.

5.4 Improvement in OCR performance with generated data

Our proposed approach was comprehensively compared to the recent state-of-the-art in Urdu handwriting generation proposed by Sharif et al. [10]. The results, presented in Table 4, indicate that training with the generated samples from our proposed approach resulted in a reduction of the Character Error Rate (CER) from 7.12 to 6.15, compared to the CER of 6.77 from the previous approach. This improvement can be attributed to the better and content-controlled generation of ligatures through our proposed approach, as opposed to the uncontrolled generation in the previous approaches, which required manual labeling or separation of ligatures. The presence of annotated samples allows for more effective and faster training and improvement of our OCR.

6 CONCLUSION

In this study, a GAN-based model was proposed to generate handwriting samples with improved quality. The model uses a pre-trained recognition network and was trained on two datasets. The model was found to produce content-controlled samples with quality comparable to recent approaches.

The proposed model also demonstrated the potential for transfer learning in handwriting generation by utilizing rendered data. While the model showed slightly higher FID scores and limited variation in ligature formation, it indicates that incorporating a larger dataset of real handwriting data with the rendered data could lead to improved results. This study lays the foundation for further research in the direction that it can be extended to generate words and sentences dynamically. Supplementing GANs with language models instead of the simple embedding layer can enhance text image generation by leveraging the combined capabilities of language understanding and receptive power to improve ligature formation.

References

1. Naeem, M. F., Awan, A. A., Shafait, F., & ul-Hasan, A: Impact of Ligature Coverage on Training Practical Urdu OCR Systems. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Vol 1, pp. 131-136, IEEE, (2017)
2. Wali, A., & Hussain, S.: Context-sensitive Shape-substitution in Nastaliq Writing System: Analysis and Formulation. In: Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, pp. 53-58, Springer Netherlands, (2007)
3. Graves, A.: Generating Sequences with Recurrent Neural Networks. In: arXiv, preprint arXiv:1308.0850, (2013)
4. Aksan, E., Pece, F., & Hilliges, O.: Deepwriting: Making Digital Ink Editable via Deep Generative Modeling. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1-14, (2018)
5. Alonso, E., Moysset, B., & Messina, R.: Adversarial Generation of Handwritten Text Images Conditioned on Sequences. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 481-486, Sydney, Australia, IEEE, (2019)
6. Ji, B., & Chen, T.: Generative Adversarial Network for Handwritten Text. In: arXiv, preprint arXiv:1907.11845., (2019)
7. Haines, T. S., Mac Aodha, O., & Brostow, G. J.: My Text in your Handwriting. ACM Transactions on Graphics (TOG), 35(3), pp. 1-18, (2016)
8. Fogel, S., Averbuch-Elor, H., Cohen, S., Mazor, S., & Litman, R.: ScrabbleGAN: Semi-supervised varying Length Handwritten Text Generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4324-4333, (2020)
9. Farooqui, F. F., Hassan, M., Younis, M. S., & Siddhu, M. K.: Offline Handwritten Urdu Word Spotting using Random Data Generation. IEEE Access, 8, pp. 131119-131136, (2020)
10. Sharif, M., Ul-Hasan, A., & Shafait, F.: Urdu Handwritten Ligature Generation using Generative Adversarial Networks (GANs). In: Frontiers in Handwriting Recognition: 18th International Conference, ICFHR 2022, Hyderabad, India, December 4–7, 2022, Proceedings, pp. 421-435, (2022)
11. El-Korashy, A. & Shafait, F.: Search Space Reduction for Holistic Ligature Recognition in Urdu Nastalique Script. In: 12th IAPR International Conference on Document Analysis and Recognition (ICDAR), Washington, DC, USA, pp. 1125–1129, (2013)
12. Riaz, N., Arbab, H., Maqsood, A., Nasir, K. B., Ul-Hasan, A., & Shafait, F.: Conv-Transformer Architecture for Unconstrained Off-LineUrdu Handwriting Recognition. In: International Journal on Document Analysis and Recognition (IJDAR) Vol 25, pp. 373–384, (2022)
13. Bin Ahmed, S., Naz, S., Swati, S., Razzak, I., Umar, A. I., & Ali Khan, A.: UCOM Offline Dataset-An Urdu Handwritten Dataset Generation. In: International Arab Journal of Information Technology (IAJIT). (2017)
14. Taylor, O.: PANGO: An Open-source Unicode Text Layout Engine. In: 25th Internationalization and Unicode Conference, Unicode Consortium, Washington DC, USA. (2004)
15. Ioffe, S., & Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: International Conference on Machine Learning, pp. 448-456, Lille, France. (2015)

16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y.: Generative Adversarial Networks. In: Communications of the ACM, 63(11), pp. 139-144, (2020)
17. Frégier, Y., & Gouray, J. B.: Mind2Mind: Transfer Learning for GANs. In: Geometric Science of Information: 5th International Conference, GSI 2021, Proceedings 5, pp. 851-859, Paris, France. (2021)
18. Zia, N. S., Naeem, M. F., Raza, S. M. K., Khan, M. M., Ul-Hasan, & A., Shafait, F.: A Convolutional Recursive Deep Architecture for Unconstrained Urdu Handwriting Recognition. In: Neural Computing and Applications, pp. 1635–1648, (2022)
19. Kingma, D. P., & Welling, M.: Auto-encoding Variational Bayes. In: arXiv, preprint arXiv:1312.6114., (2013)
20. Davis, B., Tensmeyer, C., Price, B., Wigington, C., Morse, B., & Jain, R.: Text and Style Conditioned GAN for Generation of Offline Handwriting Lines. In: arXiv, preprint arXiv:2009.00678., (2020)
21. Khrulkov, V., & Oseledets, I.: Geometry Score: A Method for Comparing Generative Adversarial Networks. In: International Conference on Machine Learning, pp. 2621-2629, Stockholm, Sweden, (2018)
22. 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. In: Advances in Neural Information Processing Systems, 30, California, USA. (2017)
23. Khattak, I. U., Siddiqi, I., Khalid, S., & Djeddi, C.: Recognition of Urdu Ligatures - A Holistic Approach. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 71-75, Washington, DC, USA. IEEE. (2015)
24. Image and Text Corpora, https://www.cle.org.pk/clestore/index.htm