

Application of deep learning on font design for ethnic minority writing systems

Yuning Zhang¹, Yikun Xi², Hongchen Chen¹, Chuhan Qiao¹,
Huiping Guo¹ and Bilige Batu¹

¹ College of Materials Science and Art Design, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia Autonomous Region, China

² Inner Mongolia Autonomous Region People's Hospital, Hohhot, Inner Mongolia Autonomous Region, China

ABSTRACT

Background: Due to the scarcity of font data, the unique writing systems of ethnic minorities are marginalized in font design and font library development, and the research and application of deep learning technologies to solve these problems face complicated challenges.

Methods: To explore font design and font library development in minority writing systems, this study used improved steps to convert font files into a unified format image files and constructed a glyph image dataset of a minority writing system. Based on a semantic skeleton of text, a method of optimizing content image selection was proposed. According to the research needs, generative adversarial networks (GANs)-based models and diffusion-based models (DMs) were preliminarily selected and pre-trained. The glyph images in the constructed dataset were input into the selected and pre-trained models to generate glyph images of the minority writing system with a new style. By comparing these two types of models, a strategy of improving efficiency and quality is proposed for glyph image generation.

Results: These efforts provide a basis for selecting image generation technology, diversifying and developing font libraries of minority writing systems, and fostering cross-cultural communication and dissemination.

Subjects Algorithms and Analysis of Algorithms, Artificial Intelligence, Computer Aided Design, Data Mining and Machine Learning, Graphics

Keywords Deep learning, Font design, Glyph image dataset, Ethnic, Writing systems, GANs, Diffusion model, AI application

INTRODUCTION

Scripts are the written form of language, used to represent the symbol system of vocabulary. The form, structure, and arrangement of words are closely related to the language they express. Different languages use different writing systems, such as Chinese characters, Latin letters, or Arabic letters. Systems include the shape of the character (glyph), their combinations, and writing direction (*e.g.*, from left to right or top to bottom).

Minority scripts are writing systems created and developed by ethnic minorities around the world. They have a long history of being used to record the language of one or more ethnic groups. Globalization and mainstream culture have put the use of many minority

Submitted 27 March 2025
Accepted 7 October 2025
Published 25 November 2025

Corresponding author
Bilige Batu, biligebatu@imau.edu.cn

Academic editor
Siddhartha Bhattacharyya

Additional Information and
Declarations can be found on
page 26

DOI 10.7717/peerj-cs.3340

© Copyright
2025 Zhang et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

writing systems at risk of decreasing or even disappearing. These writing systems often carry unique cultural, historical, and regional value that cannot be ignored when protecting glyphic heritage, promoting cross-cultural communication, or maintaining global cultural diversity. However, due to lack of digital resources and support, many ethnic minority languages have been difficult to adapt to the widespread use of information networks and rapid advancement of deep learning technologies, and their presentation and preservation on digital platforms are facing more severe challenges ([UNESCO, 2024a](#)).

UNESCO has pointed out that more than 90% of minority writing systems are not yet supported on global digital platforms ([Brookes, 2024](#)).

Even within the Unicode system, many minority scripts lack standard coding and cannot be effectively transmitted or preserved on modern digital devices, putting these writing systems at risk of extinction. On average, about 25 writing systems cease to be used each year, and many languages cannot even be fully recorded ([UNESCO, 2024b](#)).

For example, traditional scripts from parts of Africa, such as Nsibidi and Tifinagh, have largely lost their writers ([Nsibiri, 2021](#)).

The endangerment of minority languages will inevitably lead to the shrinking of multicultural development, which has a negative impact on the diversity of human social and cultural patterns, world outlook, and values. Therefore, the international community is paying more and more attention to protecting and supporting these unique writing systems through educational projects, digital protection, and policy in order to ensure their inheritance and sustainable development.

An effective way to solve the problem of language endangerment is to maintain or expand the scope of use of minority languages and make them fully used in certain regions. In the digital age, through font design and font library development, minority characters can be displayed and input on computers, mobile phones, tablets, and other devices. This not only encourages the daily use of minority languages but also helps them spread on the Internet and social media, so as to maintain vitality.

RELATED WORKS

Font design for ethnic minority writing systems

Font design is the process of creating new fonts, usually by using vector software to adjust the glyph at the dot level, which involves the creation, adjustment, and optimization of the glyph.

Developing a font library requires multiple styles of fonts. Chinese characters, Latin letters, and other scripts have font libraries with rich content, design resources, and talent support. For example, in the case of Chinese characters, artistic font generation technology and style transfer networks are updated and iterated every year. However, the research and development of fonts for ethnic minorities are extremely scarce. Ethnic minority writing systems face problems, such as a lack of font library resources and insufficient designer participation. Due to the following characteristics, there are many difficulties in font library development.

The characteristics of ethnic minority writing systems

Ethnic minority writing systems, as an integral part of a nation's cultural heritage, possess distinct features that set them apart from other writing systems. Ethnic minority writing systems mainly include the following characteristics:

- (1) Diversity: Ethnic minorities have a wide variety of writing systems, each with its own unique writing style. For example, Mongolian, Tibetan, and Uyghur scripts all have their unique letter forms and writing rules.
- (2) Historical and cultural background: These writing systems often have profound historical and cultural backgrounds, and their font designs convey language information as well as carry rich cultural connotations.
- (3) Complex glyph structure: Many ethnic minority characters are more complex in their glyph structure, which often contains a variety of strokes and decorative elements that increase the difficulty of their design and recognition.
- (4) Directionality: The writing directions of some ethnic minority writing systems are different from that of Chinese characters. Mongolian characters are written vertically from top to bottom, and Tibetan characters are written horizontally from left to right. The specific writing direction should be considered when designing font.

The difficulties in developing minority writing system fonts

Developing fonts for minority writing systems presents unique challenges stemming from their inherent characteristics and practical constraints. The difficulties in developing minority writing system fonts are mainly comprised of the following aspects:

- (1) Complex font structure: As shown in [Fig. 1](#), fonts are composed of multiple elements such as characters, glyph shapes, strokes, stroke order, stroke thickness, marks, and punctuation. When using deep learning techniques, it is necessary to consider the internal structure and proportional relationships of characters. Minority scripts have unique glyphs and structures, and similar writing systems can be distinguished by the similarities and differences of local structures to ensure that the generated fonts conform to the writing specifications of minority scripts.
- (2) Specific semantic information: Generating scripts of different fonts requires accurate reproduction of semantics, and subtle differences may lead to the generation of semantically incorrect fonts. Additionally, scripts are often closely related to culture and religious beliefs. In the development and promotion of these fonts, it is necessary to have sensitivity, respect, and understanding toward cultural connotations, and avoid cultural misuse or improper modification.
- (3) Balance between content and style: As shown in [Fig. 2](#) ([Tala, 2021](#)), line-end modeling of script strokes is the focus when forming the visual style of scripts, and it is also the main way to identify different font styles. In the design, it is necessary to control the scale of deformation in order to appropriately change the script shape according to different fonts, so that when the features are integrated into the basic content skeleton of the scripts, their recognizability and style expression can be balanced.

Exclusive punctuation mark

birga-mong	ellipsis-mong	mongolianco...	mongolianfull...	colon-mong	mongolianfou...	todosofthyph...	sibesyllablebo...	manchucomm...	manchufullsto...	nirugu-mong
?	...	,	»	:	❖	—	⊥	›	»	.

Exclusive digits mark

zero-mong	one-mong	two-mong	three-mong	four-mong	five-mong	six-mong	seven-mong	eight-mong	nine-mong
0	᠐	᠒	᠓	᠔	᠕	᠖	᠎	᠐	᠑

Part of the characters

ma-mong	la-mong	sa-mong	sha-mong	ta-mong	da-mong	cha-mong	ja-mong	ya-mong	ra-mong	wa-mong
ᠮ	ᠯ	ᠰ	ᠱ	ᠲ	ᠳ	ᠴ	ᠵ	ᠶ	ᠷ	ᠸ

Figure 1 Take the Mongolian script as an example to show the composition of a writing system. [Full-size](#)  DOI: 10.7717/peerj-cs.3340/fig-1

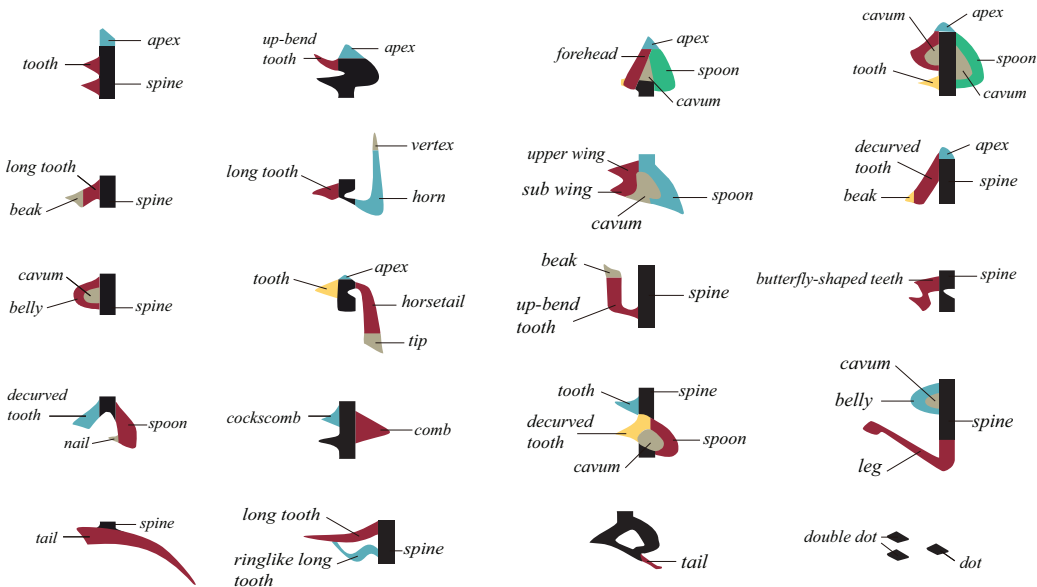


Figure 2 Line-end modeling of script strokes-taking Mongolian script as an example. [Full-size](#)  DOI: 10.7717/peerj-cs.3340/fig-2

- (4) Shortage and discontinuity of resources: The field of font design and font library development in minority writing systems faces problems such as a limited amount of reference materials, lack of long-term talent support for research, few types of fonts and uneven quality, and poor design resources and conditions.

In early font design, font generation methods primarily focused on script shape and strokes. The strokes were segmented and modeled, and the structural knowledge of the characters was used to reorganize and generate new characters ([Xu et al., 2005](#)).

Some researchers have also analyzed the style information of different fonts by constructing a database ([Zhang et al., 2011](#)), or generated scripts by extracting their font skeleton ([Velek, Liu & Nakagawa, 2001](#)).

As artificial intelligence and computer vision technology continue to develop, image generation methods based on deep learning have gradually become mainstream. In order to achieve visually realistic and semantically accurate glyph image generation, researchers have proposed a variety of models and methods. Among them, generative adversarial networks (GANs) and diffusion models (DMs) achieve more flexible glyph image generation through training and learning, promoting the rapid development of this field.

Related works of glyph image generation based on GANs

[Goodfellow et al. \(2014\)](#) was the first to propose the GAN framework. GANs use the adversarial training between two neural networks, the generator and discriminator, in order to realize the generation and simulation of complex data. In terms of image generation, the generator is responsible for generating images, and the discriminator determines the authenticity of the image. Through this confrontation, the quality and authenticity of the generated image are improved.

Later, the generator design, based on the encoder-decoder architecture combined with convolutional neural networks (CNN), made significant progress in the practical application of generating glyph images. CNN extracts features through convolutional and pooling layers to achieve efficient image recognition and classification ([Chakraborty et al., 2024](#)).

With the development of deep learning, a series of GAN-based glyph image generation networks have been proposed to continuously promote the development of this technology. A dual encoder was designed for the style and content of the image in the encoder-decoder architecture, allowing the GANs to generate high-quality content and symbol reconstruction while ensuring that the generated image accurately retains the characteristics of the symbol ([LeCun, Bengio & Hinton, 2015](#)).

[Azadi et al. \(2018\)](#) proposed an artistic character generation network based on Latin letters. The model can use six letter images to generate the remaining 20 letters. [Tian \(2017\)](#) used Namjoo Kim's ac-GAN, Taehoon Kim's dc-GAN, and Phillip Isola's original pix2pix torch code to propose a glyph image generation model for zi2zi Chinese character fonts that can complete style transfer by providing hundreds to a thousand or so glyph image samples.

[Li et al. \(2021\)](#) proposed a few-shot font style transfer method, Font-Trans GAN, which aims to solve the problem of font style transfer between different writing systems. Through

the cross-writing system style transfer model, it is possible to transfer the font style of one writing system to the font of another writing system with only a small number of source writing system font image samples. Using the style transfer experiment between different writing systems, this article verifies the effectiveness and efficiency of this method in the case of few-shot learning and provides a new solution for font design and digitization of writing systems in multilingual environments.

The style consistency of the generated fonts has attracted the attention of researchers. On the basis of Font-Trans GAN, [Zhang, Man & Sun \(2022\)](#) proposed a method of generating few-shot artistic characters in a multi-writing system called MF-Net. This method focuses on generating fonts with uniform styles for different writing systems and captures style features through few-shot learning.

As research progresses, higher demands have emerged for the information expression of generated fonts, stability of font fusion, and fidelity of font images. [Park, Hassan & Choi \(2022\)](#) proposed the CCFont model, which uses Chinese character components based on conditional GAN to generate all Chinese characters across various styles. When compared with methods that obtain all information from the image of the entire character, this model obtains more accurate information and results in less information loss, thereby reducing the failure of style conversion and producing high-quality results.

[Qin, Zhang & Zhou \(2022\)](#) proposed a font fusion network based on GANs and disentanglement representation learning to generate novel fonts, improving the interpretability of deep learning and enhancing the stability of font fusion.

[Hassan, Memon & Choi \(2023\)](#) proposed a conditional font GAN (CFGAN) that has a complex network architecture designed to generate a diverse font character set that is novel and consistent in style. This method uses the original GAN method to generate an infinite number of font styles, but focuses on real-time generation of realistic font images like photos.

Related works on generating glyph images based on DM

DM have been widely applied in various fields due to their powerful ability to generate real samples, especially in image generation where their role has become increasingly important. Their remarkable performance is not only outstanding in traditional unconditional generation, but also comparable in more complex text or image conditions.

The core principle of DM is based on a process of gradually adding noise and then gradually removing it, ultimately restoring high-quality images from the noise. Unlike GAN, DM gradually construct images through a random process, starting from a pure noise state and gradually introducing structured information until a complete image is formed. This process does not rely on discriminators, but achieves image detail restoration through the reverse process of predicting noise ([Ho, Jain & Abbeel, 2020](#)).

At present, there is not much research on generating glyph images based on DM and their variants.

For writing systems composed of a large number of complex glyph shapes (such as Chinese), [Li & Lian \(2024\)](#) introduced a method of synthesizing a small number of fonts

that can efficiently generate high-resolution glyph images. HFH-Font adopts a DM-based generation framework with component aware conditions to learn different levels of style information to adapt to different input reference sizes, thereby automatically generating large Chinese vector fonts with quality comparable to fonts manually created by professional font designers.

[Huang et al. \(2024\)](#) proposed Diff-Font, a one-time font generation framework based on DM for languages with complex ideograms and a large number of characters, such as Chinese or Korean. The model treats font generation as a conditional generation task. For scripts with rich glyphs (such as Chinese and Korean), the input of strokes or components is merged into fine conditions to generate stable training, which can learn and retain the integrity of character structure.

GANs and DMs have achieved outstanding results in the field of image generation due to their functions and characteristics in image formation. However, research on font design and font library development is still in its infancy and requires more examination of fonts in minority writing systems. Most of the results remain experimental, and their practical application in the creation of minority writing systems has not been fully verified. Therefore, this study proposes a scheme to generate fonts using artificial intelligence to solve the lack of font library resources, insufficient participation of designers, and difficulties in the sustainable development in minority writing systems:

- (1) Using experimental comparison, GAN-based models and DM models that meet the designer's design purposes were selected;
- (2) The selected models were used to generate the glyph images of minority writing systems; and
- (3) The performance and advantages of these two models in generating glyph images were compared, which provides a reference for the selection of future image generation technology.

MATERIALS AND METHODS

In order to achieve the goal of developing a proprietary font library for ethnic minority writing systems, an experiment was conducted. The experiment process included:

- (1) Constructing a glyph image dataset for ethnic minority writing systems;
- (2) Selecting models suitable for the research objectives;
- (3) Pre-training and deployment of the models; and
- (4) Generating glyph images.

All applications and research conducted in this study were performed on a workstation equipped with an NVIDIA RTX 4090 GPU (24 GB VRAM). The development environment was set up using Visual Studio Code as the primary code editor. This study strictly followed ethical norms, with specific measures for cultural sensitivities of ethnic minority writing systems: (1) A cultural advisory group of 3 experts in ethnic minority languages/cultures oversaw project design and content review, ensuring digital processing

aligned with cultural connotations and avoided misinterpretation or offense. (2) A group of 30 observers provided quantifiable scoring data for the subjective evaluation method. (3) All minority script samples came from public academic literature, official cultural classics, and authorized intangible heritage institutions. (4) Outcomes are restricted to academic exchange and cultural preservation; commercial use or dignity-harming purposes are prohibited. This project was carried out by a multi-ethnic team: Mongolian (Bilige Batu, Chen Hongchen), Daur (Guo Huiping), Manchu (Qiao Chuhan), and Han (Zhang Yuning).

Building a glyph image dataset for ethnic minority writing systems

When using GANs and DMs, a glyph image dataset for ethnic minority writing systems is essential ([Corvi et al., 2023](#)). However, high-quality image materials related to ethnic minority writing systems are scarce. Therefore, this study needed to first construct a glyph image dataset for ethnic minority writing systems.

The construction steps included: collecting open-source fonts, converting font files to glyph image files, and processing glyph images ([Järvenpää, 2023](#)).

Downloading open-source fonts

The font files required for research were downloaded from open-source platforms or websites (e.g., Google Fonts; Menksoft; onon font; ALMAS).

Converting font files to glyph image files

A font file typically contain most of glyphs from one or more writing systems; some include only common glyphs due to long design cycles or few contributors.

Since the model used here only processes image formats and cannot directly interpret font files (generally in TTF format), downloaded font files first needed conversion to glyph images ([Chen et al., 2019](#)).

The improved conversion steps were:

- (1) Import modules in Colab: Use Python code from PIL import Image, ImageDraw, ImageFont for image processing and import os for file operations ([Gujar et al., 2013](#)).
- (2) Set image parameters: Define 256-pixel size, define output path ([Naik, 2023](#)) and save as PNG.
- (3) Define character range: Obtain start/end Unicode values *via* ord(""); traverse this range with a for loop and create ImageDraw objects for rendering ([Guan, Zhou & Zhou, 2019](#)).
- (4) Configure font: Load files using ImageFont.Truetype ([pillow, 2010](#)), calculate character dimensions with draw.textsize for centering, and draw black characters *via* draw.text ([Clark, 2010](#)).
- (5) Save files: Use img.save ([Clark, 2010](#)) to store images in the specified folder, with filenames containing characters and format suffixes.
- (6) Export files: Download compressed files *via* files.download, using the make_targz_one_by_one function ([z1445056258, 2020](#)) with output filename and image folder path for compression.

Through the above steps, the codes could generate an image dataset with a specified Unicode range. This dataset was saved to a preset folder, downloaded locally, and named MEG-1.0, containing multi-style glyph images of multiple minority writing systems. The data can be retrieved from figshare (DOI 10.6084/m9.figshare.28643645).

A systematic inter-annotator validation was conducted on the dataset. An evaluation team was formed, consisting of three types of personnel: scholars specializing in ethnic minority language research, font designers with over 5 years of experience, and ordinary users.

Group glyph images

Training GAN-based models requires pre-training on multi-style glyph datasets. Since MEG-1.0 ('Converting font files to glyph image files') is insufficient, we used open Latin and Chinese glyph datasets (Li et al., 2021) to train a pre-trained model capable of cross-writing system style transfer. Data is available at https://github.com/ligoudaner377/font_translator_gan.

For inference, the model transfers style from writing system A to B. Style images were selected from MEG-1.0 and fed into the style encoder to learn style features.

Content images used high-recognition, well-structured fonts. For Mongolian, we overlaid MG-1.0 glyphs to derive a basic skeleton. Noto Sans Mongolian glyphs best matched this skeleton, so their converted images formed the content set, input to the content encoder for learning semantic skeletons.

Selecting the models suitable for the research objective

GANs and DMs, leveraging deep learning, are pivotal for glyph generation. GANs enable cross-writing system style transfer *via* few-shot learning, generating new-style glyphs from limited data. DMs use script-image multimodal input for lightweight font fine-tuning, rendering from rough to detailed images to ensure precision and diversity.

This study compared various GANs and DMs, selecting models aligned with research goals for subsequent glyph generation using deep learning.

Selecting the models suitable for the research objective from GANs

GANs are an advanced computer vision model with various types and architectures suitable for different tasks. Due to the need to ensure both the content skeleton of scripts and the style characteristics of glyphs in font generation, traditional structured GANs cannot meet the goals of this stage well. Therefore, a special structure model is needed to meet the needs of this study.

First, the selected model needed to adopt a dual encoder-decoder structure in the generator section (Li et al., 2021). Second, the discriminator of the selected model needed to include a content discriminator and a style discriminator.

In this study, MF-Net (Zhang, Man & Sun, 2022), FtransGAN (Li et al., 2021), and FCA-GAN (Zhao et al., 2024) (all open-source) with dual encoder structures were initially selected for experimental comparison. Then, based on the experiment results, the model that best met the needs of this study was selected to complete the subsequent tasks.

Selecting the models suitable for the research objective from DMs

Diffusion models, an advanced image generation technique, effectively handle glyph modeling and rendering. Their conditional generation mode enables glyph generation from specific inputs (Yang et al., 2024), ensuring user-specific needs and precise control. Multi-modal inputs (e.g., content images, text prompts) enhance detail and style adaptability (Nair, Valanarasu & Patel, 2025).

For this study, visually focused models with design flexibility—three checkpoint models (Real Vision, 2D-anime, Virtual Modeling) and three open-source low-rank adaptation (LoRA) models (WaterGlyph, Iconic, Planar) were initially selected. Checkpoints form the DM-base (RightBrain, 2022), while LoRAs enable efficient fine-tuning. After comparing the results generated by different combinations of the checkpoint and LoRA models, we selected the combination that best met the needs of this study for subsequent experiment procedures.

Pre-training and deployment of models

Training models based on GANs

First, we downloaded GitHub open-source code, set up the required environment, and organized the FtransGAN team's open Chinese-Latin font dataset (Li et al., 2021). Training hyperparameters (learning rate, batch size, optimizer) were configured (Yang & Shami, 2020) to initiate training.

Then, we started post-training. Visdom server (FOSSASIA, 2018), tracked adversarial and loop consistency losses for real-time monitoring. Training stopped when Visdom showed stable, non-fluctuating loss convergence (Hendrycks, Lee & Mazeika, 2019), indicating a well-trained model.

Deployment and parameter adjustment of models based on diffusion process

We downloaded the checkpoint and LoRA models from the open source website (Wang et al., 2023), loaded the Checkpoint model on the startup engine of the DM (Zhu, 2024), loaded the LoRA model in the prompt word and set the weight (Hu et al., 2021), and then adjusted the relevant parameters that affect the font generation effect, including sampling method (Croitoru et al., 2023), iteration step (Yang et al., 2023), the selection of ControlNet modes (Zhang, Rao & Agrawala, 2023), and prompt (Witteveen & Andrews, 2022).

The weight of this experiment was set to 0.5, with the sampling method using denoising diffusion probabilistic models (DDPM) (Ho, Jain & Abbeel, 2020). The ControlNet mode was used to select edge detection, making it suitable for tasks with fine requirements for generating boundaries.

Generating glyph images

The models selected in 'Selecting the models suitable for the research objective' (including GANs and DMs) were employed to generate glyph images of minority writing systems, thereby exploring their applicability in font design and font library development for minority languages as outlined in the research framework.

Generating glyph images using selected GAN-based models

First, we loaded the trained model, used the code to enter the test mode ([Shanmugamani, 2018](#)), and modified the file path of the content input source and style input source to MEG-1.0. Second, a set of glyph images from a minority writing system was selected from MEG-1.0 as the content basis, a set of glyph images from any writing system was selected from MEG-1.0 as the reference style, and the test command was started to make the model output the glyph images of a minority writing system with a new style. Then, index.html was used to check the generated glyph image ([Levering & Cutler, 2006](#)).

Finally, we evaluated whether the semantic content of the generated image was complete and whether the style was similar to the reference image ([Wang & Gupta, 2016](#)).

Generating glyph images using selected DM

The selected models were loaded.

Since the diffusion process can only input one content image at a time, a group of glyph images of the minority writing system were selected from MEG-1.0 and input into the model one by one to generate the corresponding images.

Evaluation criteria for generated images

Before font design, glyph images generated by models need to meet requirements at a design level ([Borgo et al., 2013](#)). The existing image evaluation criteria are mainly used to measure the difference between generated and real images ([Heusel et al., 2017](#)).

There is no relevant evaluation criteria for generated images used to make fonts when using model inference. Therefore, this study employs two evaluation methods. In addition to the commonly used objective evaluation method FID, a subjective evaluation method is added, which relies on the subjective judgment of human observers and is suitable for scenarios in this study involving visual experience, aesthetic perception, or semantic understanding.

For the subjective evaluation, 30 eligible observers (including ordinary users and professional designers) were recruited to score image quality based on their subjective perceptions. A 5-point scoring system was adopted. Observers scored the images according to preset criteria, and the image quality was reflected through statistical weighted average scores.

The targeted evaluation criteria that includes the following aspects:

- (1) Content: Check whether the glyph image restores the semantic content ([Wang et al., 2021](#)) of the text and whether there is a deviation.
- (2) Style: Evaluate whether the style of the glyph image conforms to the transferred source style ([Azadi et al., 2018](#)). Evaluate whether the style of the glyph image meets the required style and meets the designer's expectations.

- (3) Proportion: Analyze whether the proportions between the elements inside the glyph image ([Borgo et al., 2013](#)) (*i.e.*, stroke weight, letter spacing, and the height and width of the glyph) are reasonable.
- (4) Detail: Check whether the details of the glyph image are sufficient, or if the lack of details may lead to reduced recognition of the glyph content and insufficient expression of the glyph style.
- (5) Redundance: Evaluate whether there are redundant elements in the glyph image that hinder the clarity and readability of the glyph. Redundancy can appear as excessive decorative elements or unnecessary rendering effects that affect the functionality ([Neumeier, 2017](#)) of the glyphs.

According to the task requirements, we named the above standards as “Five-Dimension Subjective Glyph Quality Assessment (5d-SGQA)”. And 5d-SGQA can be integrated to comprehensively evaluate the generated glyph images.

RESULTS

In order to explore the method of font design and font library development of the minority writing system, this study first applied the improved steps to convert font files into unified-format image files and constructed a glyph image dataset of minority writing system, MEG-1.0. Second, according to the research needs, GAN-based models and diffusion-based models were preliminarily selected and pre-trained. Then, the glyph images in the constructed dataset MEG-1.0 were input into the selected and pre-trained models to generate glyph images of a minority writing system with a new style. Finally, the two methods were compared by two evaluation criteria and a strategy of improving efficiency and quality was proposed for glyph image generation.

Construction of glyph image data set of a minority writing system

We constructed a dataset of glyph images with multiple styles and writing systems that consisted of a total of 610 sets and 94,900 samples, and included glyph images from both a minority writing system and a non-minority writing system. The former was used as a content image in the study, while the latter was used as a style image. This dataset provided abundant resource support for further research and development.

Since some writing systems contain tens of thousands of glyphs, a full collection may lead to a large number of similar or repeated glyphs, which not only increases the storage cost, but also makes data processing and analysis more complicated. In deep learning, too much data may lead to an extension of training time and waste of computing resources. Therefore, this study purposefully selected representative and high-quality glyphs during conversion and deleted blank font images caused by the lack of original font files after conversion ([Table 1](#)).

Selecting the models suitable for the purpose of the study

Selecting the GAN-based model suitable for research purposes

As shown in [Table 2](#), the experiment took the generation of Mongolian font images as an example. MF-net, Ftrans-GAN, and FCA-GAN were used to transfer the style of Chinese

Table 1 Built glyph image dataset including multiple styles and writing systems (MEG-1.0).

Script category		Number of fonts	The number of selected glyphs	The final number of samples (delete some blank glyphs)
Minority	Armenian	68	77	5,196
	Bengali	25	91	2,232
	Devanagari	67	94	6,246
	Miao	23	146	3,342
	Tibetan	42	191	7,417
	Uighur Mongolian	45	148	2,654
	Uighur Mongolian	45	148	2,654
Non-minority	Arabic	57	256	6,408
	Chinese	44	993	39,720
	Hiragana	38	83	3,143
	Katakana	38	89	3,311
	Hangul syllable	20	324	6,480
	Korean alphabet	37	51	1,887
	Thai alphabet	62	74	4,576
	Latin	44	52	2,288

character images to Mongolian fonts. Each pre-trained model was tested 300 times, and then the generated results tended to be stable.

Table 2 shows that MF-net performed best in the style transfer task, the reason being that the model adopted an efficient multi-scale feature extraction and cross-domain mapping strategy. This kind of model can usually maintain the content of the original image well, and capture and apply the unique style characteristics of the scripts through deep learning.

The content of the images generated by Ftrans-GAN was biased, and the shape structure of the scripts was not as stable as MF-net. The fluency of strokes and maintenance of details were slightly rough. In particular, the generated results showed a certain distortion for the subtle changes and local structure of the strokes.

The style transfer stability of FCA-GAN was poor and the generated images had style deviation or distortion. In particular, the clarity and accuracy of the generated images was inferior to the previous two when handling complex calligraphy or strokes.

Selecting the diffusion-based model for research purposes

As shown in Table 3, we used combinations of different checkpoint models and LoRA models, and the input content image was controlled by text prompt words to generate a new style of glyph image.

The Real Vision Checkpoint model could maintain the clarity of the overall contour and structure of the glyph during the generation process. For the stroke ratio and structure of standard glyphs, the results of the Real Vision Checkpoint were closer to the standards of real fonts, especially in detail processing, showing higher accuracy. Due to the more stable and easily adjustable generated results, it proved particularly suitable for professional font development and design, and the workload of subsequent processing was relatively small.

Table 2 Comparison of the images generated by GANs-based models. The style of Ground truth is the style reference sample of Mongolian character designed by font designer according to the Chinese character style image.

Category	Times			
	1st	100th	200th	300th
Style				
Content				
MF-net				
Ftrans-GAN				
FCA-GAN				
Ground truth*				

Note:

*The Ground truth images are designed by the font designers in our research team and the fourth author, Huiping Guo.

The glyph images generated by the 2D-anime Checkpoint model contained more unnecessary light and shadow effects, which may be because the model is biased towards the characteristics of a 2D animation style, resulting in too many decorative elements in the generated glyph. The overall structure of the glyph had a large change in the generation process, and the length and thickness ratio of the strokes was not consistent with the original content. It was suitable for the font style generation of creative design or animation effects, but not suitable for scenes that require strict glyph standards.

The generated results of the Virtual Modeling Checkpoint were clear, but the control of details was inferior to the Real Vision Checkpoint model. The glyph generated by the model performed well on the overall contour, had a stable structure, and maintained a high consistency with the input content. In terms of detail processing, such as the thickness of strokes and fluency of starting and ending strokes, the model's performance was slightly insufficient and not as detailed as Real Vision Checkpoint.

Based on its overall performance, WaterGlyph LoRA had a stable generation effect, with uniform stroke contours and character details, and no obvious deformation or discontinuity. The overall structure of the character shape was well maintained, and the generated stroke connections were smooth and the details were uniform, indicating that it can better preserve the characteristics of the input content. The stability and accuracy of WaterGlyph LoRA make it highly suitable for font designs that require high levels of detail and contour.

The glyph generated by Iconic LoRA may appear stiff or show a loss of fluidity at bends, making it difficult to meet the delicate requirements of standard glyph shapes. Although the overall outline maintained a certain degree of integrity, the results deviated from

Table 3 Comparison of the images generated by the combination of diffusion-based models. The style of Ground truth is the style reference sample of Mongolian character designed by font designer according to the Chinese character style image.

Category	Original	WaterGlyph LoRA	iconic LoRA	planar LoRA
Content				
Real vision checkpoint				
2D-anime checkpoint				
Virtual modeling checkpoint				
Ground truth				

expectations in terms of details, especially in the handling of curved strokes and starting and ending strokes. It may have advantages when dealing with relatively simple stroke structures, but its effectiveness is limited for complex glyph design requirements.

The processing of the start and end points of glyph strokes by Planar LoRA was slightly insufficient, and the generated glyph appeared slightly rough and not smooth in its end-point representation. Although the overall outline of the character shape was acceptable, there was a gap in the accuracy and consistency of details compared to WaterGlyph LoRA.

Comparing GAN-based models and diffusion-based models

The GAN-based model and the diffusion-based model used in the experiment were different in structure (Figs. 3 & 4). In the process of generating images, their emphasis is also different. The former focuses on style transfer, and the latter focuses on repair and reconstruction.

Comparing the structure of GAN-based models and diffusion-based models

GANs and DMs are definitely different in structure (Figs. 3 & 4). In addition to their structural differences, GANs optimize image quality through the confrontation process of

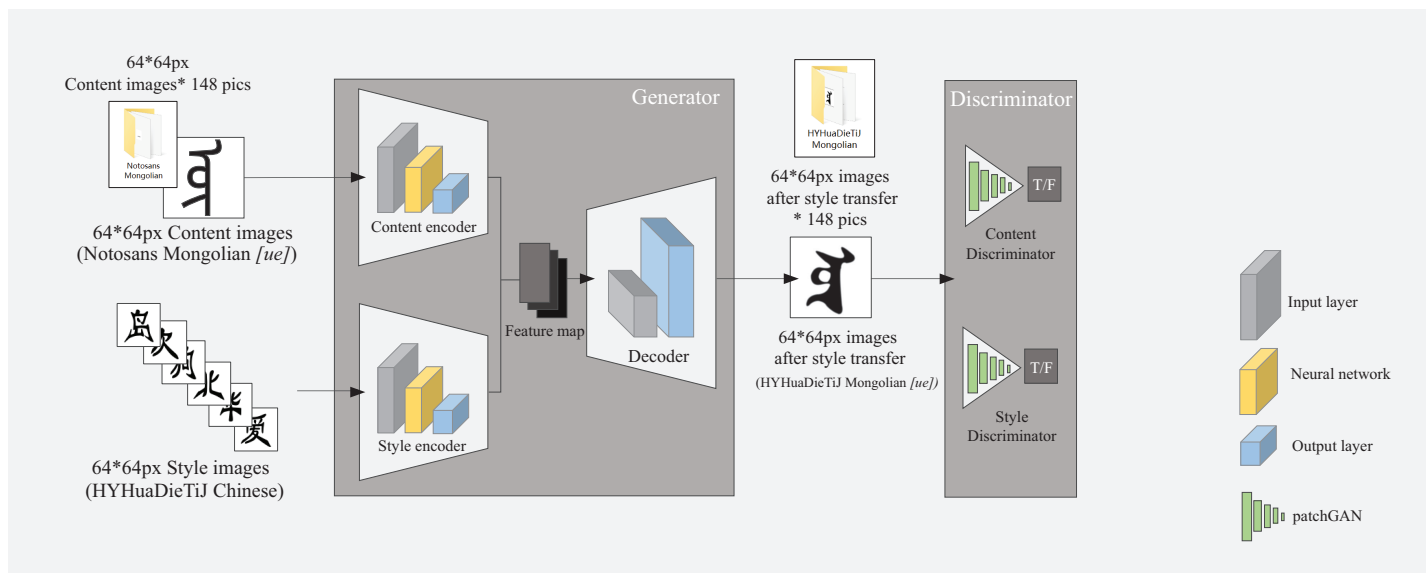


Figure 3 Structure of cross-writing system style transfer model based on GAN.

Full-size DOI: 10.7717/peerj-cs.3340/fig-3

generators and discriminators, but they are usually limited to a single mode and rely more on existing image data for generation, while the DMs can accept multiple input modes. Therefore, users can jointly guide the generation of glyph images through text description and image reference, and gradually generate clear glyph images through the noise reduction process.

The following is a comparison of the type of input image, generation process, and generated result when generating a glyph image using the GANs and DMs.

Comparing the types of input images of GAN-based models and diffusion-based models

GANs usually generate images from random noise vectors and accept specific conditional inputs (images or text) to control the generated image style. Additionally, the input is generally a fixed dimension. As shown in Fig. 3, 148 content and style images of 64 * 64 px were entered in this study, all of which were black characters with white background.

DMs can start directly from the existing image, gradually add noise to the image and deduce it in reverse to generate a clear image. It can also generate clear images through multi-modal inputs, such as text and initial images. As shown in Fig. 4, inputting one glyph image and style image of any size and inputting text control generated clearer images in this study.

Comparing the generation process of GAN-based models and diffusion-based models

The GAN models (Fig. 3) are trained by a generator and a discriminator for adversarial training. This process generates all glyph images at once without the need for multiple iterations and reverse generation, resulting in a faster generation speed.

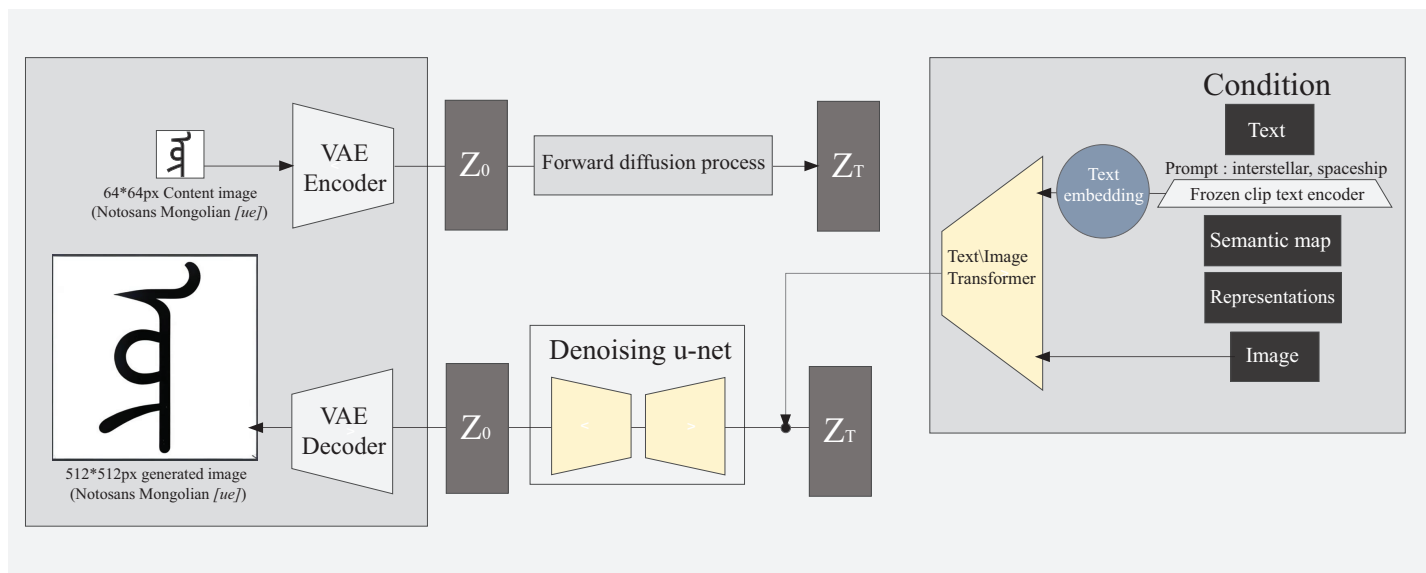


Figure 4 Structure of fine-tuned multimodal image generation model based on diffusion model. Full-size [DOI: 10.7717/peerj-cs.3340/fig-4](https://doi.org/10.7717/peerj-cs.3340/fig-4)

The generation process of the DMs is gradual (Fig. 4), starting from a completely denoised image and gradually restoring the target image through a series of denoising steps. Each step of generation requires calculating the result of the previous step, which means that only one glyph image can be generated at a time. Therefore, the generation process of the DM is relatively slow, but the quality and details of the generation can be more finely controlled.

Comparing the generated results of GAN-based models and diffusion-based models

The MF-net and the combination of Real Vision Checkpoint and Water Glyph LoRA, selected in 'Selecting the models suitable for the research objective', were used to generate glyph images. From the output results, three glyph images were chosen to compare generation effects.

As shown in Table 4, after successful training, GANs can quickly generate images, and is especially proficient at generating realistic images in a short time, but the images generated multiple times lacked diversity. However, the DM can optimize details through subsequent multiple generation, which is suitable for generating high-quality and detailed images. Although the generation speed is slow, the generated images are more diverse and detailed, which is especially suitable for tasks that need to control multi-modal input.

In this study, the advantages of a GAN-based models are its fast generation speed, strong learning ability of style features, and suitability for large-scale image generation and transfer generation of style features. The diffusion-based model pays more attention to the details and clarity of generated images. Although its generation speed is

slow and the style characteristics are not as detailed, it performed better in fine control and stability.

Evaluating GAN-based models and diffusion-based models

In this study, the weights are defined based on the comprehensive consideration of the importance of each element and the needs of the evaluation scenario. Finally, content and style are set as core elements (0.25 each), proportion and detail as important dimensions (0.20 each), and redundancy as a less influential factor (0.10). The data used are determined by collecting feedback from a questionnaire survey involving 30 recruited participants.

Evaluation results using FID

Figures 5 and 6 visualize FID scores of different models across epochs. For GAN-based models in Fig. 5, lower FID means better performance. MF-Net shows the best stability with steadily low FID. Ftrans-GAN initially has high FID but improves, while FCA-GAN has consistently high FID, indicating weaker performance.

In Fig. 6, Real Vision Checkpoint achieves the lowest FID at the WaterGlyph LoRA stage, indicating optimal performance here. Its FID rises later but remains competitive. Virtual Modeling Checkpoint shows a rising FID trend, implying reduced efficacy.

Evaluation results using 5d-SGQA

To comprehensively assess model performance *via* 5d-SGQA, two radar charts visualize subjective scores across five dimensions (Style, Content, Proportion, Detail, Redundance).

Figure 7 focuses on single GAN-based models (FCA-GAN, Ftrans-GAN, MF-net). MF-net outperforms others in most dimensions, showcasing strong semantic restoration, style transfer, and detail expression. Ftrans-GAN follows with balanced scores but lags behind MF-net. In contrast, FCA-GAN has significantly lower scores across all dimensions, indicating weaknesses in multi-dimensional performance.

Figure 8 evaluates diffusion models. The combination of Real Vision Checkpoint and Water Glyph LoRA achieves the highest scores in key dimensions (Content, Style, Detail), demonstrating excellent semantic accuracy, style consistency, and detail richness, with well-controlled redundancies.

DISCUSSION

The study first converted limited minority font files into glyph images to construct the MEG-1.0 dataset. Then, it inputted the dataset into multiple GAN models and diffusion models to generate new-style minority glyphs, and compared the results of these models. Finally, based on the comparison, it selected a generation workflow combining GAN (MF-Net) and diffusion models (Real Vision Checkpoint with WaterGlyph LoRA).

Contrast experiments were conducted for three reasons:

- (1) Minority scripts face resource scarcity and structural complexity; comparing models verifies their suitability for these unique scripts, addressing data inadequacy.

Table 4 Generated images by GANs-based model and diffusion-based model. Content images were input at one time, including all the glyphs of one writing system, take the Mongolian script as an example, 148 images will be generated in 4s (NVidia 4090). Only one content image can be input at a time, and it takes roughly 14s to output an image.

Model	Content	Output		
	0 time	1st time	50th time	300th time
MF-net				
	0 time	1st time	2nd time	3rd time
Real vision checkpoint & Water Glyph LoRA				
	0 time	1st time	2nd time	3rd time

- For designers, it evaluates whether generated images meet content accuracy and style expression needs, exploring possibilities to optimize image performance and methods.
- Technically, single models have limitations (GANs generate quickly but lack diversity; diffusion models excel in quality but are slow); comparison clarifies their advantages, providing a basis for integrating strengths.

Enhancing efficiency and quality and optimizing font image generation

The writing systems of ethnic minorities face the dilemma of resource scarcity and difficulty in inheritance. Designing and producing a mature and usable font requires an experienced design team and a considerable amount of time to complete.

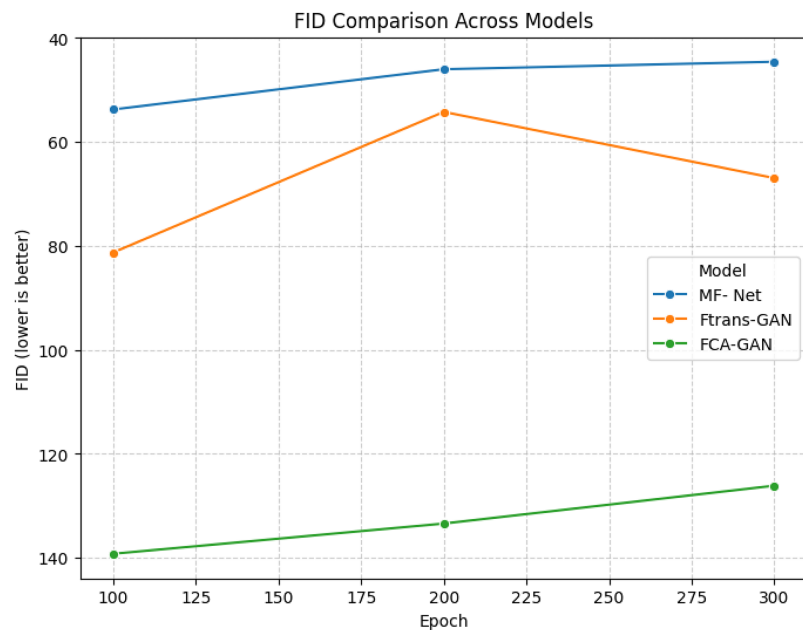


Figure 5 Data visualization line chart of GANs' generation results evaluated using FID.

Full-size DOI: 10.7717/peerj-cs.3340/fig-5

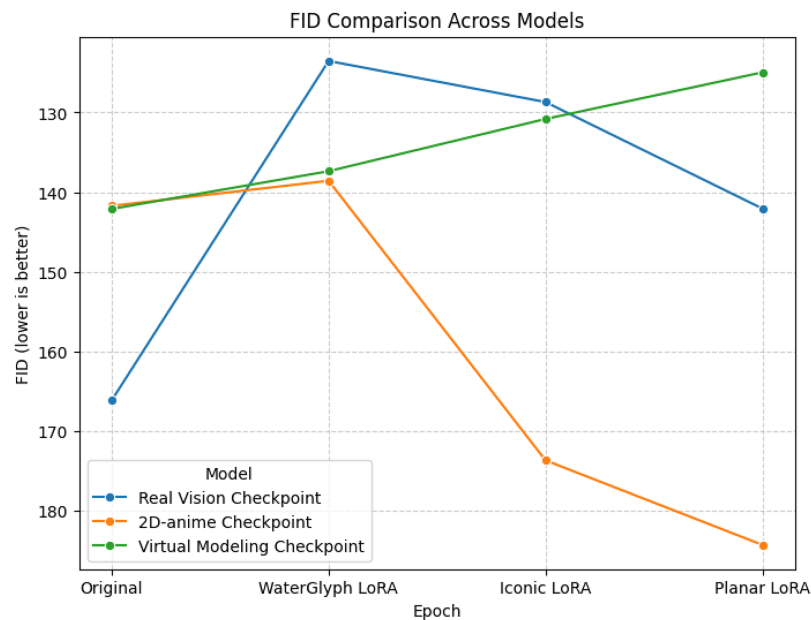


Figure 6 Data visualization line chart of DMs' generation results evaluated using FID.

Full-size DOI: 10.7717/peerj-cs.3340/fig-6

In the field of deep learning, in order to rectify the difficulties in font production, some researchers have proposed GAN-based models to generate glyph images with new font styles, which have performed well in the generation of Chinese and Korean characters (Tian, 2017). Subsequently, other researchers introduced CNN and encoder-decoder

Radar Chart of Average Scores for GANs



Figure 7 Radar chart of average scores for GANs.

Full-size DOI: 10.7717/peerj-cs.3340/fig-7

Radar Chart of Average Scores for DMs

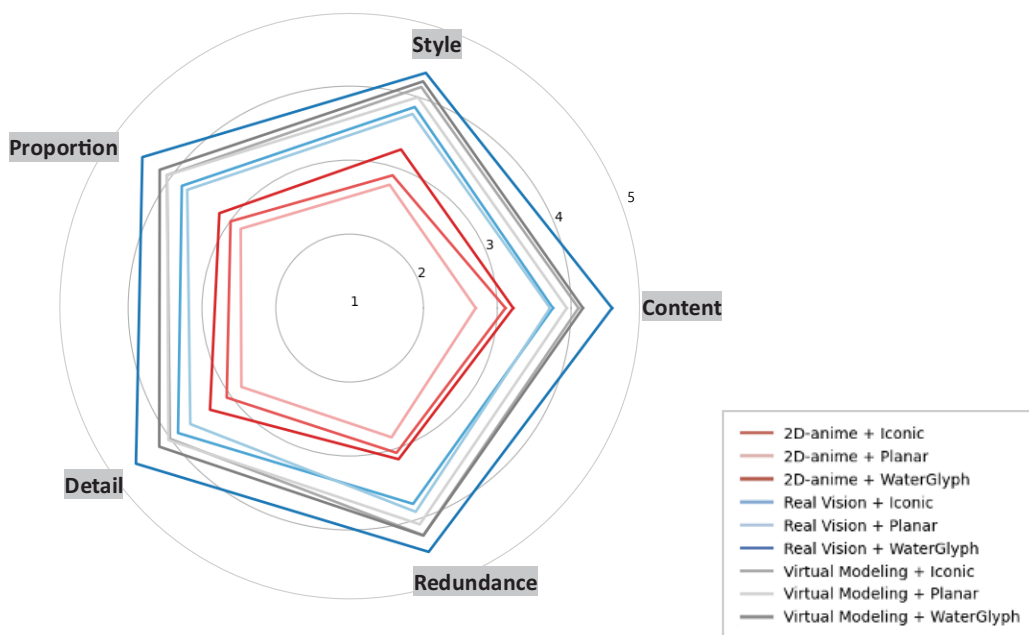


Figure 8 Radar chart of average scores for DMs.

Full-size DOI: 10.7717/peerj-cs.3340/fig-8

structures into GANs to achieve style transfer across writing systems (Li et al., 2021) and few-shot learning (Zhang, Man & Sun, 2022), which performed well in the style transfer of Latin letters and Chinese characters. However, during our experimental process, it was found that the generated new style glyphs were far from the expected goals in this study, when these GAN-based models were applied to the writing system of ethnic minorities.

After the development of DMs, many technology companies encapsulated them as image generation artificial intelligence (AI) and provided them to users around the world for use. Many designers attempted to use these AI to generate new glyph images, but the actual process of generation is to make the text three-dimensional and then perform special-effect rendering (Recraft, 2022), rather than generating pure glyph images with new font styles. Meanwhile, researchers in academia mainly focused on how to enable DMs to generate semantically accurate English words directly within the image while simplifying input conditions (Yang et al., 2024). However, the font style of the generated words is often limited to common and well-established digital fonts.

Although DMs can generate high-definition images, its generation speed is slow and it can only accept a single input command at a time, which limits its application field. In recent years, researchers have attempted to combine DM with GAN to accelerate the inference generation process of images by increasing the sampling speed of the model (Trinh & Hamagami, 2024; Xiao et al., 2024), or by combining the advantages of GANs and DMs in a multimodal manner to improve image quality and generation efficiency in generation tasks (Yenew, Assefa & Belay, 2024). The models combined with these studies are a fixed structure with strong pertinence.

There are some difficulties in making the image quality generated by GANs reach that of the image generated by DM, or making the efficiency of DMs generation catch up with that of GANs in the field of font image generation and font production. There is still a lack of a feasible and flexible solutions for stably and efficiently generating high-quality images, especially for minority writing systems.

Therefore, this study proposes a two-step strategy: select GANs for batch generation, then DMs for optimization. Preliminary experiments chose MF-net and Real Vision Checkpoint with WaterGlyph LoRA. MF-net enabled one-time style transfer for full script glyphs; DM optimized these to enhance style and clarity.

As shown in Table 5, three glyph images were chosen from the glyph images generated by MF-net, and the model combination of Real Vision Checkpoint and WaterGlyph LoRA was used to optimize these three glyph images. The glyph images with more obvious style features and clearer contours were obtained.

The obtained glyph images not only can be used to create digital fonts, but also directly as data to expand the glyph image dataset of ethnic minority writing systems so that researchers can train models with better performance. At the same time, in the field of minority writing system datasets, due to insufficient datasets during model training, the goal of few-shot learning could not be achieved. This method efficiently and effectively expands the dataset, thereby achieving the training of a few-shot learning model and further improving the efficiency of font generation.

Table 5 The glyph images generated by the combination of GAN and diffusion models.

Input	Output		
Generated by MF-net	Optimized by real vision checkpoint model & Water Glyph LoRA		
			

Collecting method of glyph images of minority writing system

When collecting glyph images of writing systems as datasets, researchers generally use two methods. One is to manually collect images with text and then use data annotation to classify images. This method is very costly and time-consuming. The other is to use the code font2img ([Li et al., 2021](#); [Zhang, Man & Sun, 2022](#)), developed by professionals, to batch-convert font file in TTF format into image format.

However, in this study, it was found that due to the confusion of font file format, lack of unified design standards, and severe version breakage of font files in ethnic minority writing systems, the text position in the source font was not at the center point, the font size was different in different versions, and there was blank Unicode, making it impossible to directly collect glyph images from writing systems using the above methods.

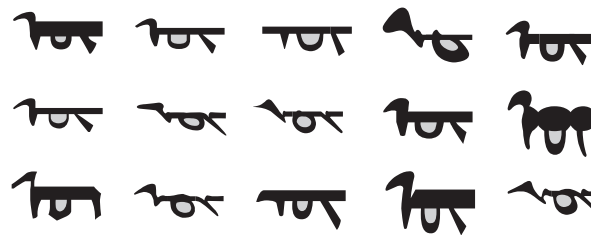
Therefore, this study proposes an effective collection method to solve these problems (see ‘Converting font files to glyph image files’ for details), including: using font editing software FontCreator to check the Unicode encoding in the font file, confirming the character intervals defined by the encoding, recording the position and size of the text, rewriting the traversal range of commands, calculating the center point, and adjusting the text size for each writing system to ensure that the glyph images converted from the font file are valid.

By this method, this study constructed its own dataset and named it MEG-1.0.

Optimizing content images based on semantic skeleton

When using pre-trained GAN-based models for style transfer, researchers ([Li et al., 2021](#); [Zhang, Man & Sun, 2022](#); [Zhao et al., 2024](#)) generally do not define or select content images, but randomly choose a set of glyph images from the dataset as content image to input models.

Mongolian script element *[ue]*'s representation of glyphs in different font files



[ue] in Notosans Mongolian font file



Overlap multiple *[ue]* glyph images

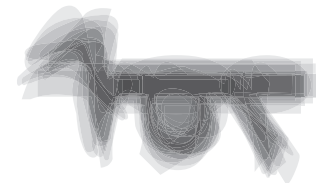


Figure 9 The reason for using NOTOSANS series as a content image set (taking Mongolian characters as an example). [Full-size](#) DOI: 10.7717/peerj-cs.3340/fig-9

The purpose of the content encoder is to learn the skeleton of the text, and the glyph image randomly selected as the content image may come from the font with low semantic value. Based on this phenomenon, when selecting content images, this study obtained the font whose glyph outline best aligned with the semantic skeleton of the text through repeated test (Fig. 9). The most suitable font for the content image was the Noto Sans font series.

Establishing a results-oriented new evaluation method

In practical application tests, this study found that for font design tasks, due to the outstanding performance of DMs in image restoration and creative generation, the selection of generated images depends not only on objective evaluation data but also more on subjective judgment. This is because the style generated by DMs varies significantly each time.

Therefore, this study proposed a new Subjective Assessment method (Five-Dimension Subjective Glyph Quality Assessment, 5d-SGQA) and recruited observers to evaluate the results using this method.

The designer team involved in this study believes that the FID has limited reference value. While the similarity between generated images and original images can serve as one of the evaluation dimensions, the innovativeness of generated images may be more critical. Based on this, this study assigned weights to each evaluation dimension before conducting data analysis.

The evaluation results after the weight distribution are more reasonable and effective.

Limitations and future works

The generative logic of AI models relies on patterns within training data. As observed in GAN experiments, insufficient training on minority data often leads to failures in accurately reconstructing the content skeleton of minority glyphs, highlighting potential bias.

Another limitation of this study is that its testing is primarily based on Mongolian script, with insufficient validation across other ethnic minority writing systems. The research team has ensured the accuracy of Mongolian-related results through rigorous verification.

However, Mongolian was selected as a representative case of ethnic minority writing systems based on technical principles and glyph image characteristics. Thus, the observed patterns and workflow are theoretically transferable.

It is hoped that experts specializing in other ethnic minority languages will adopt the explored application procedures to conduct their own experiments.

Given these limitations, future work will focus on addressing the identified gaps to enhance the generalizability and robustness of the proposed framework.

CONCLUSION

In response to the demand for converting font files of ethnic minority writing systems into glyph images, this study presents effective conversion steps and constructs a multi-style, multi-script dataset to support further research. It also proposes a semantic skeleton-based method to optimize content image selection, enhancing glyph generation quality.

Given GANs' subpar output quality, DMs' slow generation and weak style, and rigid GAN-DM combinations, this study develops a strategy to rapidly expand high-quality datasets. By selecting suitable GANs, DMs, or combinations, it efficiently generates high-quality glyphs, resolving data scarcity in minority scripts. Adopting technical exploration to address script use and digitization issues, this study features a universally applicable technical workflow and tool selection logic.

This study's font generation scheme has substantial practical potential. Extending it to 10 major ethnic minority languages would reach ~1 million core users (average 100,000 per language; data from The Collection of Chinese Language Resources: Ethnic Language Volume, 2021).

For real-world font users, such as wayfinding systems, designers previously spent time hand-drawing or customizing minority fonts individually; our solution enables quick output of scenario-adaptive minority scripts based on mainstream ones. In cultural brand visual design, constrained by scarce high-quality font resources, the MEG-1.0 dataset provides direct references for designers.

In field research, our solution offers efficient tools for cultural protectors, aiding endangered script rescue and inheritance. It provides methods for interdisciplinary field research, lowering technical barriers—simplifying conversion of non-standard script samples into analyzable data for ethnographic and anthropological researchers.

ACKNOWLEDGEMENTS

Figure 2 in this study was produced with the help of its original author, Tala. Thanks to them for their efforts in this study.

Thanks to all the research teams mentioned in the references for the open source project of glyph style migration as well as the open data sets. Thanks to YUODAO Translator

(which provides translation services based on artificial neural networks) and peerj's professional polishing team.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by the 2023 Inner Mongolia Autonomous Region Key Research and Development and Achievement Transformation Project entitled "Research and Development of Automatic Generation Technology for Uyghur-Mongolian Digital Fonts" (Project No.: 2023YFSW0004). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

2023 Inner Mongolia Autonomous Region Key Research and Development and Achievement Transformation Project: 2023YFSW0004.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yuning Zhang conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Yikun Xi performed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Hongchen Chen analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Chuhan Qiao performed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Huiping Guo analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Bilige Batu conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available at Figshare: ZHANG, YUNING (2025). MEG-1.0.zip. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.28643645.v1>.

The code is available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.3340#supplemental-information>.

REFERENCES

- Azadi S, Fisher M, Kim VG, Wang Z, Shechtman E, Darrell T. 2018.** Multi-content GAN for few-shot font style transfer. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 7564–7573.
- Borgo R, Kehrer J, Chung DH, Maguire E, Laramee RS, Hauser H, Ward M, Chen M. 2013.** Glyph-based visualization: foundations, design guidelines, techniques and applications. In: *Eurographics (State of the Art Reports)*, 39–63.
- Brookes. 2024.** *Atlas of endangered alphabets*. London: Quercus.
- Chakraborty T, KS UR, Naik SM, Panja M, Manvitha B. 2024.** Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. *Machine Learning: Science and Technology* 5(1):11001 DOI 10.1088/2632-2153/ad1f77.
- Chen J, Ji Y, Chen H, Xu X. 2019.** Learning one-to-many stylised Chinese character transformation and generation by generative adversarial networks. *IET Image Processing* 13(14):2680–2686 DOI 10.1049/iet-ipr.2019.0009.
- Clark JA. 2010.** Python imaging library. Available at <https://pypi.org/project/pillow/> (accessed 2 January 2025).
- Corvi R, Cozzolino D, Poggi G, Nagano K, Verdoliva L. 2023.** Intriguing properties of synthetic images: from generative adversarial networks to diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 973–982.
- Croitoru F-A, Hondru V, Ionescu RT, Shah M. 2023.** Diffusion models in vision: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(9):10850–10869 DOI 10.1109/TPAMI.2023.3261988.
- FOSSASIA. 2018.** A flexible tool for creating, organizing, and sharing visualizations of live, rich data. Supports Torch and Numpy. GitHub. Available at <https://github.com/fossasia/visdom> (accessed 2024).
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio YJA. 2014.** Generative adversarial nets. In: *NIPS'14: Proceedings of the 28th International Conference on Neural Information Processing Systems*, Vol. 27.
- Guan Y, Zhou F, Zhou J. 2019.** Research and practice of image processing based on python. In: *Journal of Physics: Conference Series*. Bristol, IOP Publishing, 22018.
- Gujar MH, Mhatre MP, Ghanate MS, Kadam MS, Kurle MD, Shitole S. 2013.** Python based image processing. Available at https://www.researchgate.net/publication/309208697_Python_Based_Image_Processing.
- Hassan AU, Memon I, Choi J. 2023.** Real-time high quality font generation with conditional font GAN. *Expert Systems with Applications* 213:118907 DOI 10.1016/j.eswa.2022.118907.
- Hendrycks D, Lee K, Mazeika M. 2019.** Using pre-training can improve model robustness and uncertainty. In: *International Conference on Machine Learning*. Westminister: PMLR, 2712–2721.
- Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter SJA. 2017.** Gans trained by a two time-scale update rule converge to a local Nash equilibrium. In: *Neural Information Processing Systems (NIPS)*, Vol. 30.
- Ho J, Jain A, Abbeel PJA. 2020.** Denoising diffusion probabilistic models. In: *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vol. 33, 6840–6851.
- Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, Wang L, Chen W. 2021.** LoRa: low-rank adaptation of large language models.

- Huang Y, Huang J, Liu Y, Yan M, Lv J, Liu J, Xiong W, Zhang H, Chen S, Cao LJ. 2024. Diffusion model-based image editing: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 47:4409 DOI 10.1109/TPAMI.2025.3541625.
- Järvenpää K. 2023. Automatic font parameter optimisation. Available at https://www.researchgate.net/publication/376424562_Automatic_Font_Parameter_Optimisation (accessed 1 November 2025).
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521(7553):436–444 DOI 10.1038/nature14539.
- Levering R, Cutler M. 2006. The portrait of a common HTML web page. In: *Proceedings of the 2006 ACM Symposium on Document Engineering*. New York: ACM, 198–204.
- Li H, Lian Z. 2024. HFH-Font: few-shot Chinese font synthesis with higher quality, faster speed, and higher resolution. *ACM Transactions on Graphics* 43(6):1–16 DOI 10.1145/3687994.
- Li C, Taniguchi Y, Lu M, Konomi Si. 2021. Few-shot font style transfer between different languages. In: *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Piscataway: IEEE, 433–442 DOI 10.1109/WACV48630.2021.00048.
- Naik PG. 2023. *Conceptualizing Python in Google Colab: hands-on practical sessions*. Bilaspur, Chhattisgarh: Shashwat Publication.
- Nair NG, Valanarasu JMJ, Patel VM. 2025. Maxfusion: plug&play multi-modal generation in text-to-image diffusion models. In: *European Conference on Computer Vision*. Springer, 93–110.
- Neumeier J. 2017. Legibility in typeface design for screen interfaces. Master’s thesis, Aalto University School of Arts, Design and Architecture, Aalto, Finland. Available at <https://underscoretype.com/static/files/neumeier-johannes-legibility-in-typeface-design-for-screen-interfaces.pdf>.
- Nsibiri. 2021. Writing in Africa—modernising nsibidi. Available at <https://thelanguagecloset.com/2021/01/23/writing-in-africa-modernising-nsibidi/>.
- Park J, Hassan AU, Choi J. 2022. CCFont: component-based Chinese font generation model using generative adversarial networks (GANs). *Applied Sciences* 12(16):8005 DOI 10.3390/app12168005.
- pillow. 2010. ImageFont module. Available at <https://pillow.readthedocs.io/en/stable/reference/ImageFont.html#PIL.ImageFont.truetype> (accessed 1 November 2025).
- Qin M, Zhang Z, Zhou X. 2022. Disentangled representation learning GANs for generalized and stable font fusion network. *IET Image Processing* 16(2):393–406 DOI 10.1049/ipr2.12355.
- Recraft. 2022. Premium image generation and editing tool. Available at <https://www.recraft.ai/about>.
- RightBrain. 2022. vega.ai. Available at <http://applet.rightbrainai.cn/> (accessed 2024).
- Shanmugamani R. 2018. *Deep learning for computer vision: expert techniques to train advanced neural networks using tensorflow and keras*. Birmingham, UK: Packt Publishing Ltd.
- Tala. 2021. A brief discussion on the method of shape splitting and naming of Uighur-Mongolian characters. *Journal of Inner Mongolia Agricultural University (Social Science Edition)* 13(3):198–199+220.
- Tian Y. 2017. Master Chinese calligraphy with conditional adversarial networks. GitHub. Available at <https://github.com/kaonashi-tyc/zi2zi> (accessed 1 November 2025).
- Trinh LT, Hamagami T. 2024. Latent denoising diffusion GAN: faster sampling, higher image quality. *IEEE Access* 12:78161 DOI 10.1109/ACCESS.2024.3406535.

- UNESCO. 2024a. Digital preservation of indigenous languages: at the intersection of technology and culture. Available at <https://www.unesco.org/en/articles/digital-preservation-indigenous-languages-intersection-technology-and-culture>.
- UNESCO. 2024b. The International Year of Indigenous Languages: mobilizing the international community to preserve, revitalize and promote indigenous languages. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000379771>.
- Velek O, Liu C-L, Nakagawa M. 2001. Generating realistic kanji character images from on-line patterns. In: *Proceedings of Sixth International Conference on Document Analysis and Recognition*. Piscataway: IEEE, 556–560.
- Wang C, Duan Z, Liu B, Zou X, Chen C, Jia K, Huang JJ. 2023. PAI-diffusion: constructing and serving a family of open Chinese diffusion models for text-to-image synthesis on the cloud. Arxiv DOI 10.48550/arXiv.2309.05534.
- Wang X, Gupta A. 2016. Generative image modeling using style and structure adversarial networks. In: *European Conference on Computer Vision*. Springer, 318–335.
- Wang J, Pan G, Sun D, Zhang J. 2021. Chinese character inpainting with contextual semantic constraints. In: *Proceedings of the 29th ACM International Conference on Multimedia*. New York: ACM, 1829–1837.
- Witteveen S, Andrews MJ. 2022. Investigating prompt engineering in diffusion models. ArXiv DOI 10.48550/arXiv.2211.15462.
- Xiao H, Wang X, Wang J, Cai J-Y, Deng J-H, Yan J-K, Tang Y-D. 2024. Single image super-resolution with denoising diffusion GANs. *Scientific Reports* 14(1):4272 DOI 10.1038/s41598-024-52370-3.
- Xu S, Lau FC, Cheung WK, Pan Y. 2005. Automatic generation of artistic Chinese calligraphy. *IEEE Intelligent Systems* 20(3):32–39 DOI 10.1109/MIS.2005.41.
- Yang Y, Gui D, Yuan Y, Liang W, Ding H, Hu H, Chen K. 2024. Glyphcontrol: glyph conditional control for visual text generation. ArXiv DOI 10.48550/arXiv.2305.18259.
- Yang L, Shami AJN. 2020. On hyperparameter optimization of machine learning algorithms. *Theory and Practice* 415(1):295–316 DOI 10.1016/j.neucom.2020.07.061.
- Yang L, Zhang Z, Song Y, Hong S, Xu R, Zhao Y, Zhang W, Cui B, Yang M-H. 2023. Diffusion models: a comprehensive survey of methods and applications. *ACM Computing Surveys* 56(4):1–39 DOI 10.1145/3626235.
- Yenew AB, Assefa BG, Belay EG. 2024. HouseGanDi: a hybrid approach to strike a balance of sampling time and diversity in floorplan generation. *IEEE Access* 12:125235 DOI 10.1109/ACCESS.2024.3451406.
- z1445056258. 2020. Google colab Download folder (packaged file download). Available at <https://blog.csdn.net/z1445056258/article/details/111395009> (accessed 18 December 2020).
- Zhang Y, Man J, Sun P. 2022. MF-Net: a novel few-shot stylized multilingual font generation method. In: *MM '22: Proceedings of the 30th ACM International Conference on Multimedia*. New York: ACM, 2088–2096 DOI 10.1145/3503161.3548414.
- Zhang J, Mao G, Lin H, Yu J, Zhou C. 2011. Outline font generating from images of ancient Chinese calligraphy. In: *Transactions on Edutainment*. Germany: Springer, 122–131.
- Zhang L, Rao A, Agrawala M. 2023. Adding conditional control to text-to-image diffusion models. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway: IEEE, 3836–3847.

Zhao H, Rosin PL, Lai Y-K, Meng W, Wang YN. 2024. Cross-lingual font style transfer with full-domain convolutional attention. *Pattern Recognition* **155(2)**:110709
DOI [10.1016/j.patcog.2024.110709](https://doi.org/10.1016/j.patcog.2024.110709).

Zhu A. 2024. *Using stable diffusion with Python: leverage Python to control and automate high-quality AI image generation using stable diffusion*. Birmingham, England: Packt Publishing Ltd.