GlyphGenius: Unleashing the Potential of AIGC in Chinese Character Learning

- IEEE ACCESS –

- Supplementary Material –

Linhui Wang Yi Lou Xinyue Li Yuxuan Xiang Tianyi Jiang Yiying Che
Chen Ye *

Tongji University

* yechen@tongji.edu.cn

This document includes additional results and experimental details for the main paper.

1 Ablative Study Detail

Unless otherwise clarified, we use 9.0 as a default CFG scale, 0.85 as default Denoising strength, DPM++2S a Karras as the sampler, 20 steps to sample each image. Since Stable Diffusion relies on CFG-Scale to generate high-quality images, and CFG-Scale uses a "negative prompt" to guide the denoising during inference, we use an empty string as the negative prompt in the "No Prompt" setting, while using "ugly, low-quality" in other settings. We use the four prompt settings for experiments:

No Prompt Use empty strings as input prompt, *e.g.*, "".

Insufficient Prompt prompts that do not fully cover objects in conditioning images, *e.g.*, "high quality and detailed masterpiece" that does not mention the actual image contents.

Conflicting Prompt prompts that change the semantics of conditioning images, *e.g.*, "a dog" for a cat image.

Perfect Prompt prompts that describe all necessary content semantics, *e.g.*, "a photograph of the billowing, soft colorful ribbon, 4K, sharp focus".

For comparison, we trained Stable Diffusion[1] V1.5 with LoRA[2] with exactly the same method and on the same dataset as Multi-stage Model, upon which the model used in section "Stable Diffusion V1.5 & LoRA & Writing Sequences" expands.

Figures 1 to 4 depict simple Chinese characters with fewer strokes, which still maintain strong pictographic features and possess significant potential for visual representation. Therefore, a direct visual transformation is applied to the entire Chinese character.

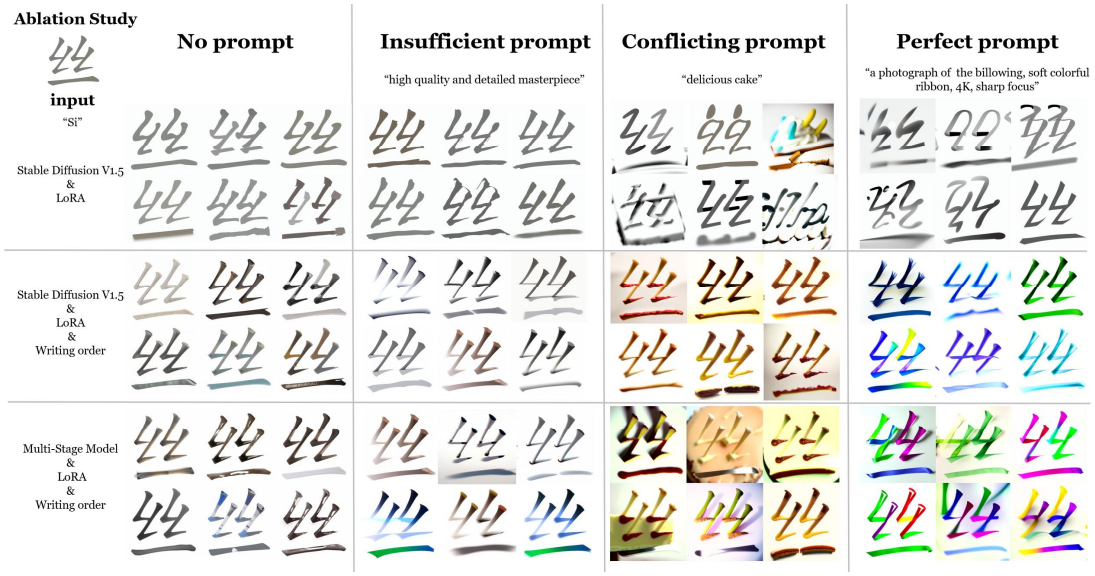


Figure 1: The result of character "丝", meaning silk or ribbons in English.

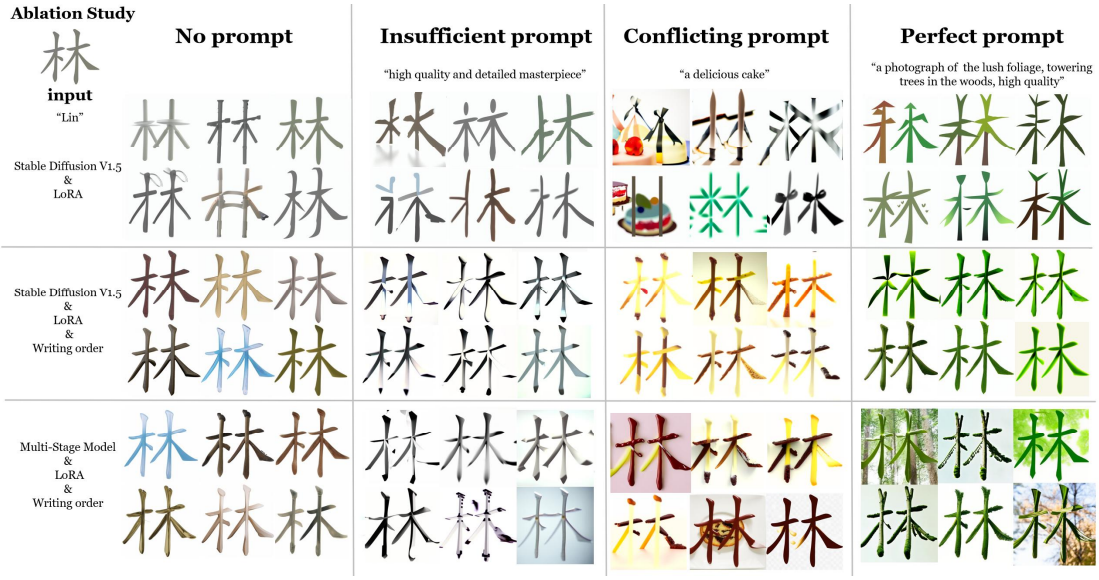


Figure 2: The result of character "林", meaning moods in English. It is composed of two character "木", which means trees in English.

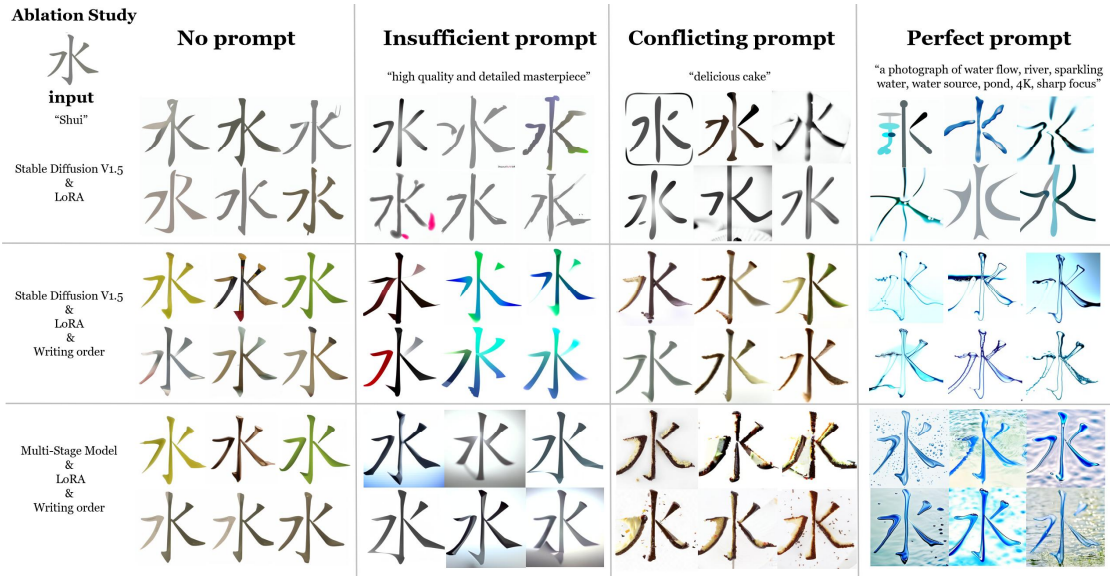


Figure 3: The result of character "水", meaning water in English.

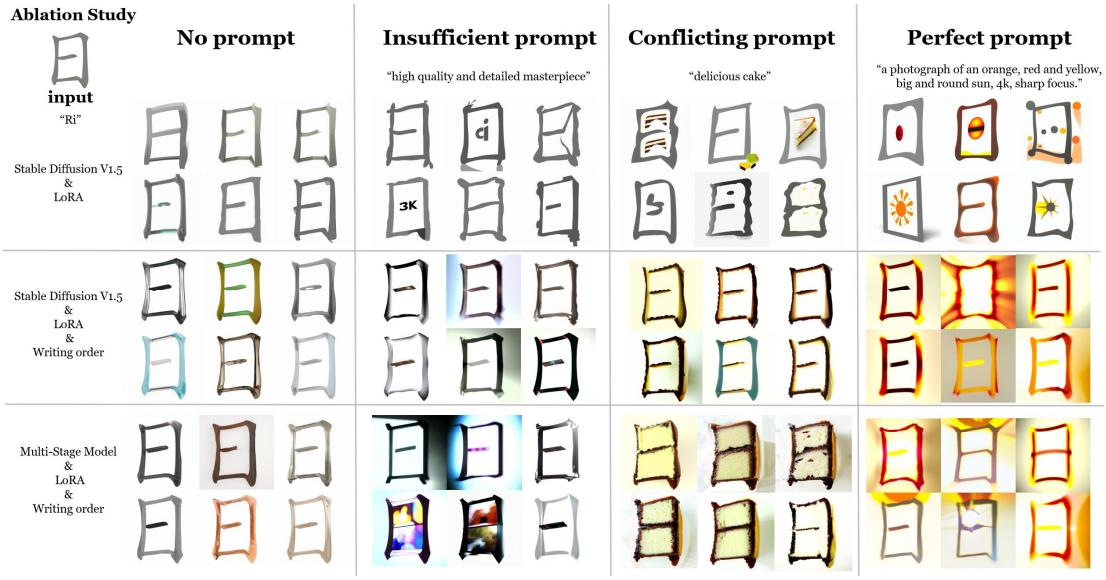


Figure 4: The result of character "日", meaning the sun in English.

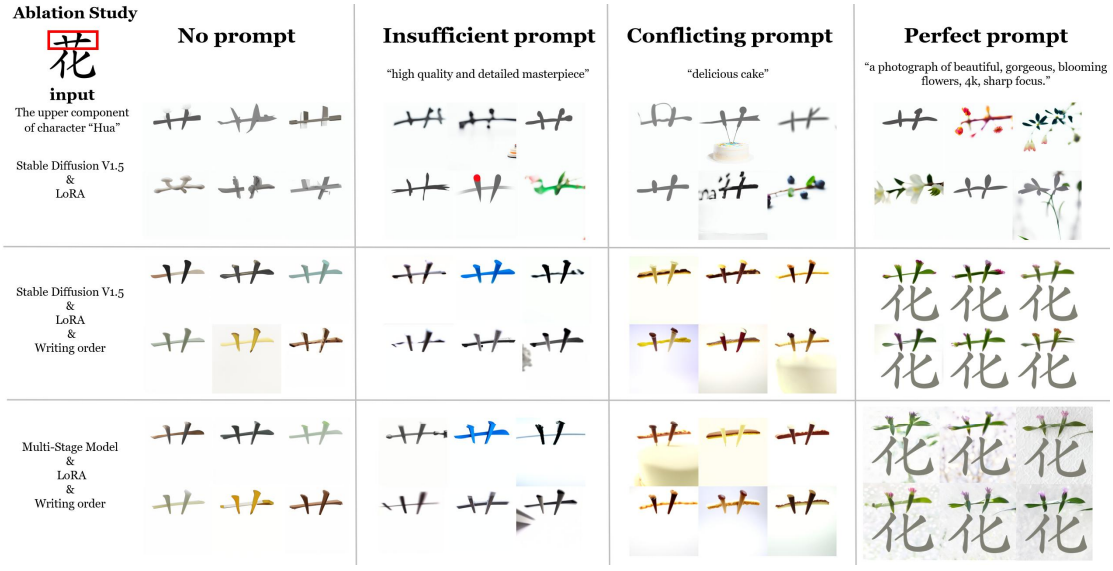


Figure 5: The result of the character "花". This character means flowers. However, its upper component means grass while lower component has irrelevant meaning.

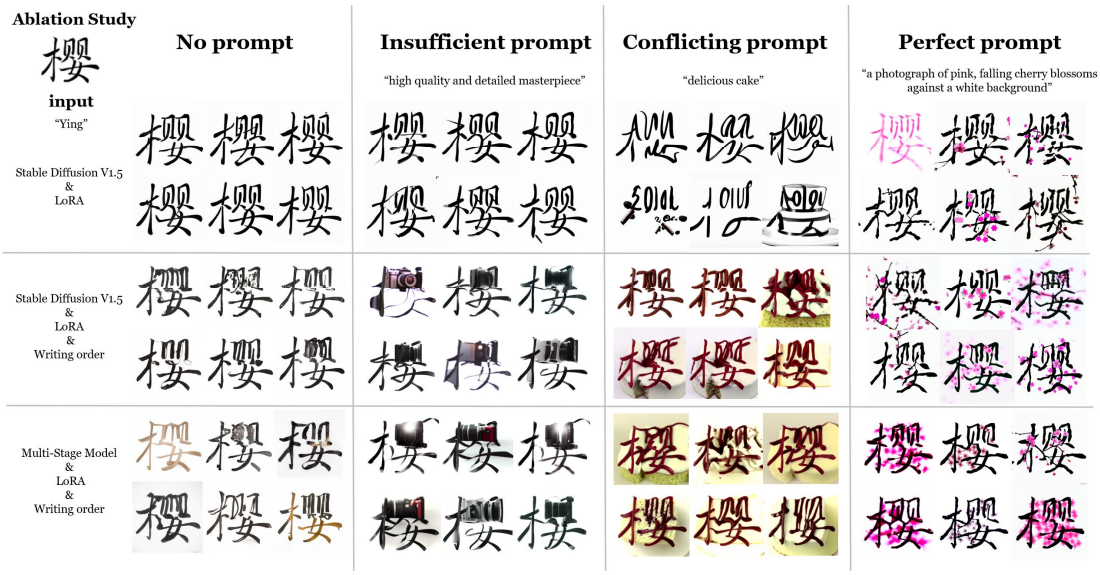


Figure 6: The result of character "樱", meaning cherry blossoms in English.

Figure 5 illustrate Chinese characters with a "top-bottom" structure. Typically, in this type of Chinese character, only one part carries pictographic meaning closely related to the overall meaning of the character, while the other part is associated with the pronunciation of the character. Hence, for visual transformation, only the components with pictographic meaning are selected. In addition, we combine it with the remaining untreated part in the column of "Perfect Prompt", to compare the visual effects of different models applied to this system's modular redraw. The first three columns remain unchanged.

Figure 6 represents complex Chinese characters with a higher number of strokes, serving as a test for evaluating the model's capability in handling intricate Chinese characters.

2 Skeleton Similarity Evaluation

In this phase, we conducted experiments on 10 representative Chinese characters. By default, we utilized a CFG scale of 9.0, denoising strength of 0.85, and DPM++2S as the sampler, with 20 sampling steps per image. All images were formatted as 256*256 pixels. For each character, we provided 4 text descriptions, generating 25 images per description. We applied a skeleton extraction algorithm[3] to extract the character's skeleton and calculated the similarity between the skeleton of each image and the original character's skeleton using three image similarity metrics: (1)SSIM [4], (2)PSNR, and (3)MSE. The final result for each character and metric was obtained by averaging the values of 100 images relative to the original character image.

To evaluate the efficacy of our multi-stage model in handling Chinese character strokes across different stages, we documented the outcomes of stable diffusion V1.5 with LoRA (SD), stage-I, and stage-II. Figures 7 to 10 depict the processing of selected Chinese characters, while Table 1 presents the processing results data for the 10 characters alongside the corresponding text descriptions used.

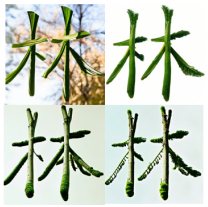
Original	Input	Skeleton	SSIM↑	RSNR↑	MSE↓
			-	-	-
Stable Diffusion V1.5 with LoRA			0.827	11.106	5059.688
Stage-I (Ours)			0.867	**12.125**	**4171.295**
Stage-II (Ours)			**0.868**	11.468	5037.165

Figure 7: The process of skeleton extraction and similarity comparison of the character “林”.

Original	Input	Skeleton	SSIM↑	RSNR↑	MSE↓
			-	-	-
Stable Diffusion V1.5 with LoRA			0.613	11.173	5000.029
Stage-I (Ours)			0.876	11.618	4507.053
Stage-II (Ours)			**0.893**	**12.377**	**3833.836**

Figure 8: The process of skeleton extraction and similarity comparison of the character “心”.

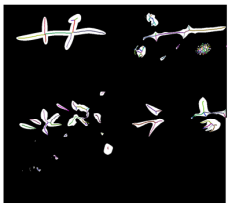
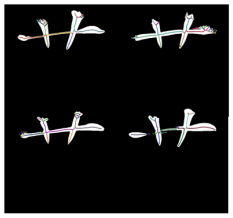
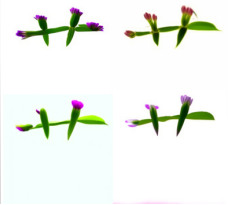
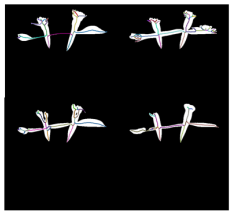
Original  The upper component of character “Hua”	Input 	Skeleton 	SSIM↑ -	RSNR↑ -	MSE↓ -
Stable Diffusion V1.5 with LoRA			0.722	14.435	2625.237
Stage-I (Ours)			**0.954**	**17.609**	**1182.168**
Stage-II (Ours)			0.948	16.922	1375.772

Figure 9: The process of skeleton extraction and similarity comparison of the upper component of the character “花”.

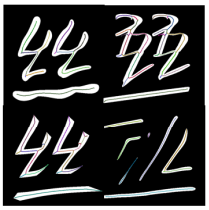
Original	Input 	Skeleton 	SSIM↑ -	RSNR↑ -	MSE↓ -
Stable Diffusion V1.5 with LoRA			0.738	11.173	5332.129
Stage-I (Ours)			**0.867**	**11.895**	**4887.307**
Stage-II (Ours)			0.863	11.382	4940.327

Figure 10: The process of skeleton extraction and similarity comparison of the character “丝”.

Table 1 Partial Results of evaluation for skeleton similarity of tow stages. We report SSIM, PSNR, MSE for our method of multi-stage generation in contrast to original stable diffusion pre-trained model.

	SSIM↑			PSNR↑			MSE↓		
	SD	Stage-I	Stage-II	SD	Stage-I	Stage-II	SD	Stage-I	Stage-II
火	0.786	**0.917**	**0.917**	12.730	**14.970**	14.498	3550.137	**2150.103**	2322.065
林	0.827	**0.842**	0.823	11.106	**12.125**	11.468	5059.688	**4171.295**	5037.165
丝	0.738	**0.867**	0.863	11.254	**11.895**	11.382	5332.129	**4887.307**	4940.827
心	0.613	0.876	**0.893**	11.173	11.618	**12.377**	5000.029	4507.053	**3833.836**
花	0.722	**0.954**	0.948	14.435	**17.609**	16.922	2625.327	**1182.168**	1375.772
樱	0.647	0.658	**0.677**	6.859	7.048	**7.335**	13429.216	12848.778	**12201.791**
日	**0.894**	0.878	0.874	**13.195**	12.590	12.129	5332.129	4887.307	**4740.827**
猫	0.863	0.876	**0.893**	11.173	11.618	**12.377**	6707.053	**5000.029**	5833.836
水	0.871	**0.933**	0.826	**11.685**	10.169	10.173	**2625.327**	3468.261	3750.232
京	0.882	**0.907**	0.893	13.476	15.225	**15.468**	3024.643	**2171.643**	2537.512

[1] Text descriptions of the character "火": (1)A roaring bonfire burning brightly.(2) A match striking out flames.(3) A burning candle.(4)A volcanic eruption.
[2] Text descriptions of the character "林": (1) Lush Green Forest.(2) Ancient Trees.(3) A pine tree with sturdy trunks and lush foliage.(4) Two trees with many branches and green leaves.
[3] Text descriptions of the character "丝": (1) Smooth and soft silk.(2) floating ribbons.(3)Silk Road.(4)colorful ribbons that are being waved.
[4] Text descriptions of the character "心" are all "A red heart symbol".
[5] Text descriptions of the character "花":(1) A blooming Peony.(2) Falling Petals.(3) blossoming flowers with green leaves.(4)A vast expanse of blooming flowers.
[6] Text descriptions of the character "樱": (1) A blossoming cherry tree. (2) Pink, falling cherry blossoms. Another two descriptions are the same as the former two.
[7] Text descriptions of the character "日" are all "an orange, red and yellow, big and round sun."
[8] Text descriptions of the character "猫": (1) A cat stretching lazily. (2) A cat pouncing in flight. (3) A cat sitting crouched on the ground with both feet. (4) A cat lying down asleep.
[9] Text descriptions of the character "水": (1) Flowing water. (2) A flowing river. (3) The shimmering surface of a light blue lake. (4) Translucent and gleaming water droplets.
[10] Text descriptions of the character "京": (1) The ancient city wall constructed with bricks.(2)A magnificent palace.(3) The bustling urban skyscraper.(4) The ornate overhanging roof eaves of a palace

3 User Study Detail

3.1 Online Guess-Meaning Quiz

Each online "Guess-Meaning" questionnaire contains a total of seven questions. Each question corresponds to a Chinese character, and the participants are required to choose one option from the given four options that they believe is correct.

To minimize the potential impact on participants' recognition in the control group, we take characters in regular script font for the control group (we refer to as "original group"). On the other hand, we take the redrawn results of the system as the experimental group (we refer to as "wordarts group"). Therefore, we can ensure that difference in recognizing a character in the experimental group is solely due to its visual form. For instance, in Figure 9, we have redrawn the left component of the character "Lang" (wolf) and combined it with the right half of the original character to create the experimental group of this character.

A total of 135 valid responses are received. We just keep 121 samples of beginner Chinese learners, who are new to Chinese language learning (within 1 year).

Figure 11~Figure 17 depict the chosen characters and corresponding question results.

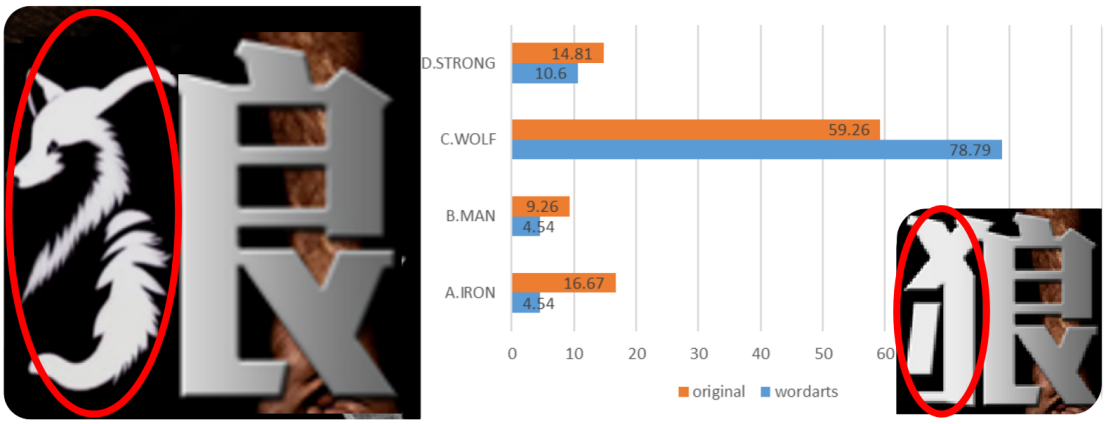


Figure 11: The result of “Lang” (means wolf in English) in this quiz.

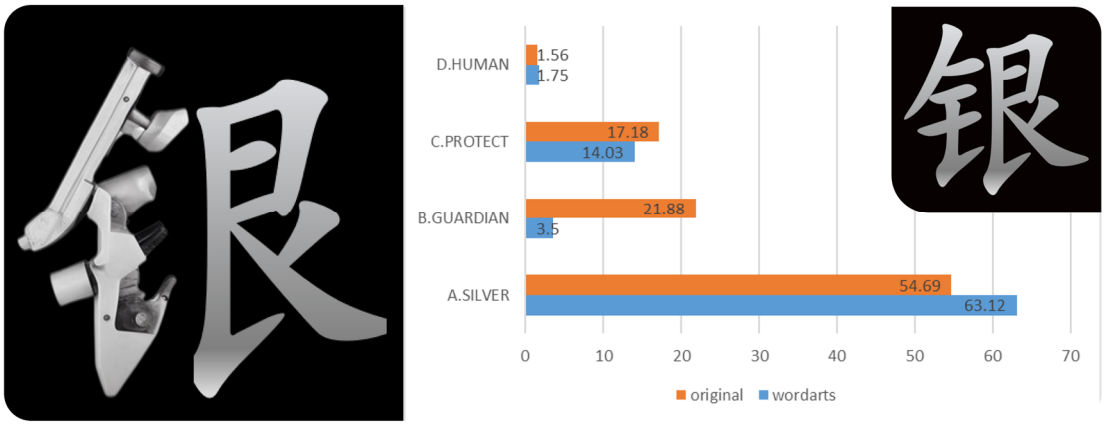


Figure 12: The result of “Ying” (means silver in English) in this quiz.

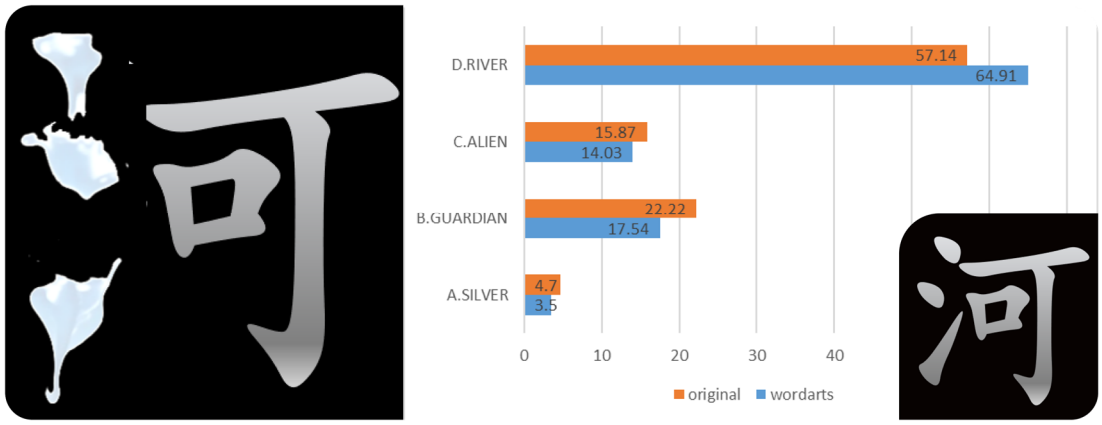


Figure 13: The result of “He” (means river in English) in this quiz.

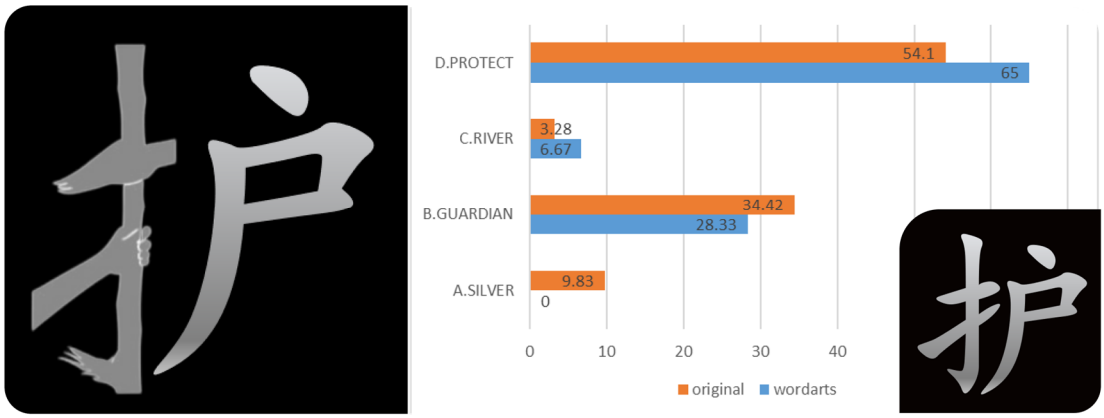


Figure 14: The result of “Hu” (means protect in English) in this quiz.

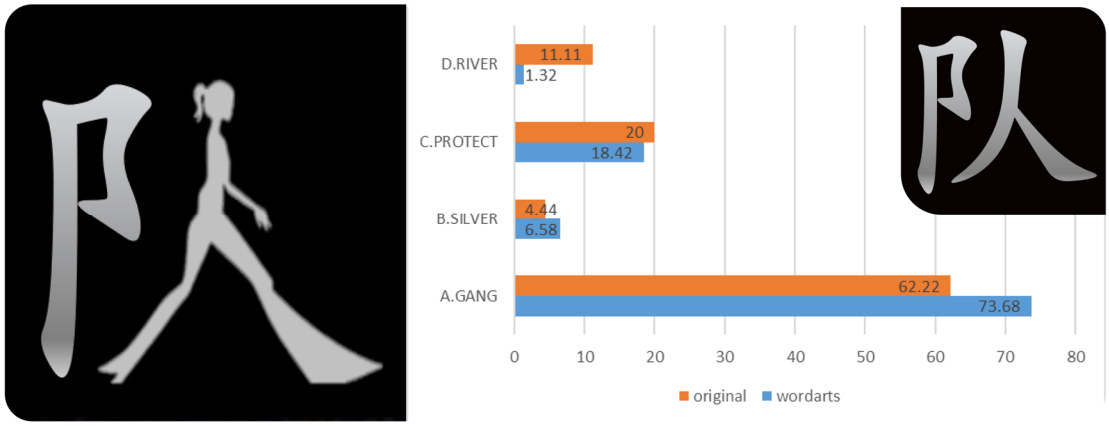


Figure 15: The result of “Dui” (means a group of people in English) in this quiz.

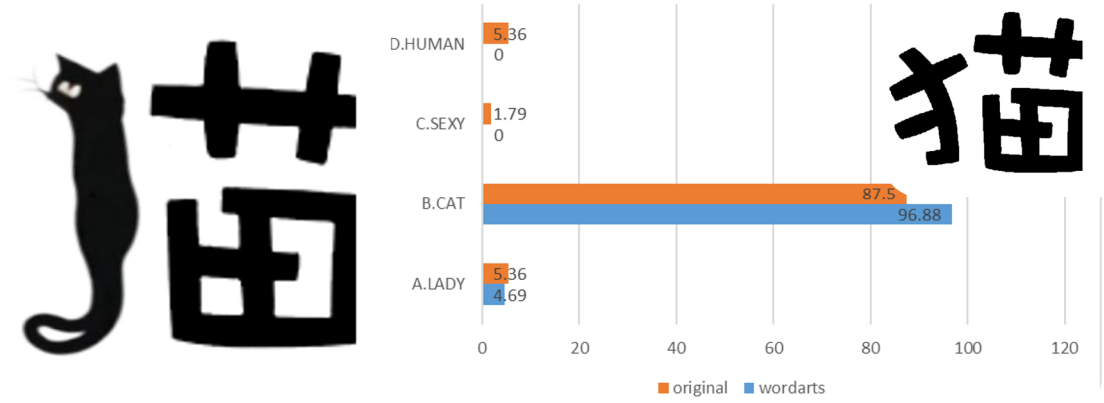


Figure 16: The result of “Mao” (means cat in English) in this quiz.

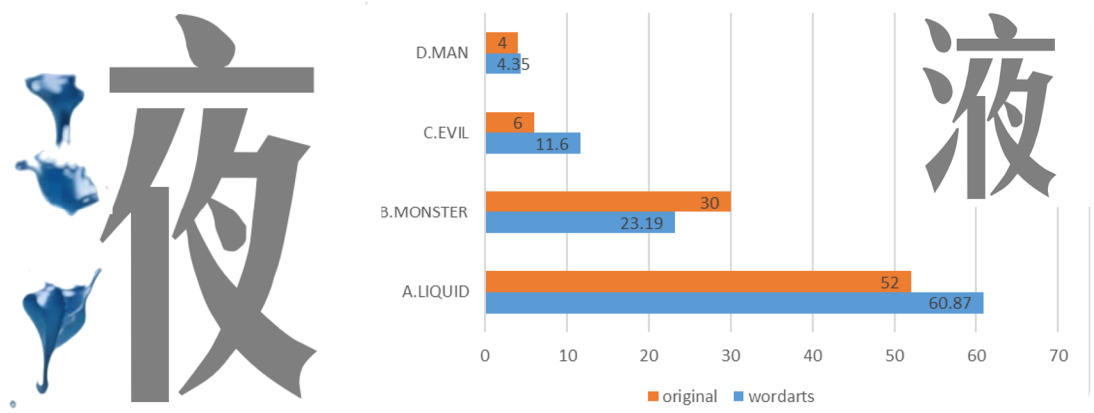


Figure 17: The result of "Ye" (means liquid in English) in this quiz.

3.2 Offline System Test

We recruited 21 volunteers (10 males and 3 females, aged between 24 and 38 with an average age of 32) and organized an offline workshop. Figure 17 shows the overview of the workshop and some redrawn results by participants.

Figure 18: Overview of the workshop. (a)Lecture of our system. (b)Design "wordarts" on the touchscreen freely with our prompt card or own ideas. (c)Fill in the questionnaire.

The questionnaire consisted of seven questions in total:

Q1 [workflow] Is the whole workflow of "Draw, Render, Reorganize" obvious and natural to you?

Q2 [performance] How satisfied you are with the output of this system?

Q3 [design] Do you think you can design creative word arts with your own ideas?

Q4 [cognition] Do you think this system is helpful for the understanding of the original meaning of Chinese character components?

Q5 [non-expert use] Do you think this system is friendly to those with little experience in graphic design?

Q6 [pedagogy] Do you think this system helps junior students or international learners in Chinese and Chinese character learning?

Q7 [overall] We expect your overall rating of our system!

Figure 19~21 below show the results of the questionnaire.

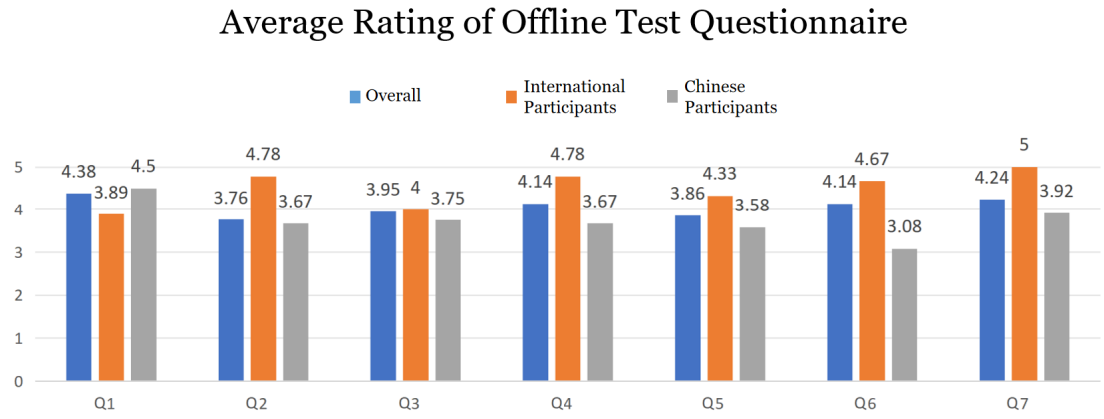


Figure 19: The overall average score of each question for overall participants, international participants and Chinese participants.

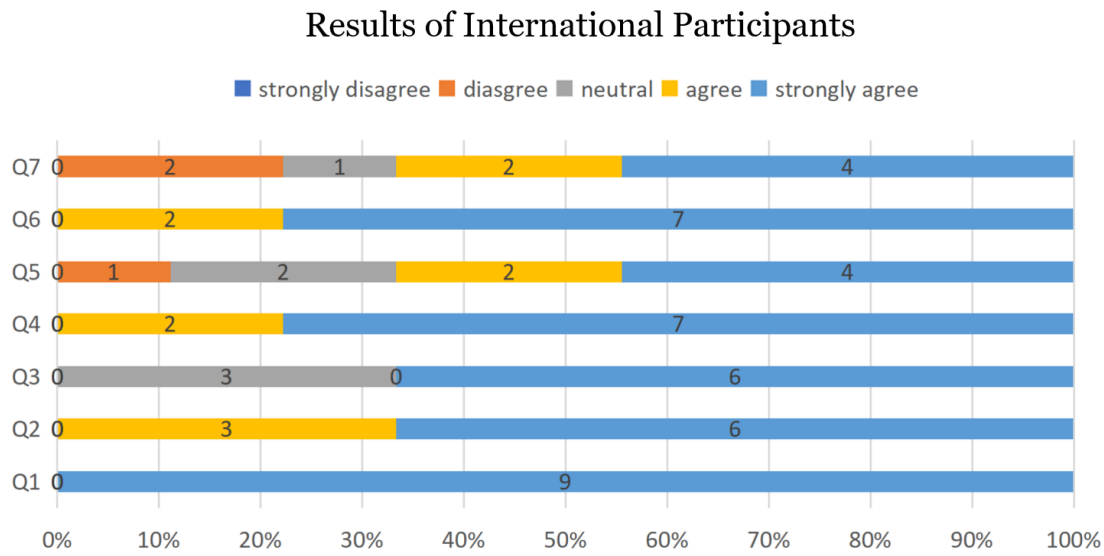


Figure 20: The distribution of scores assigned by international participants across the seven questions.

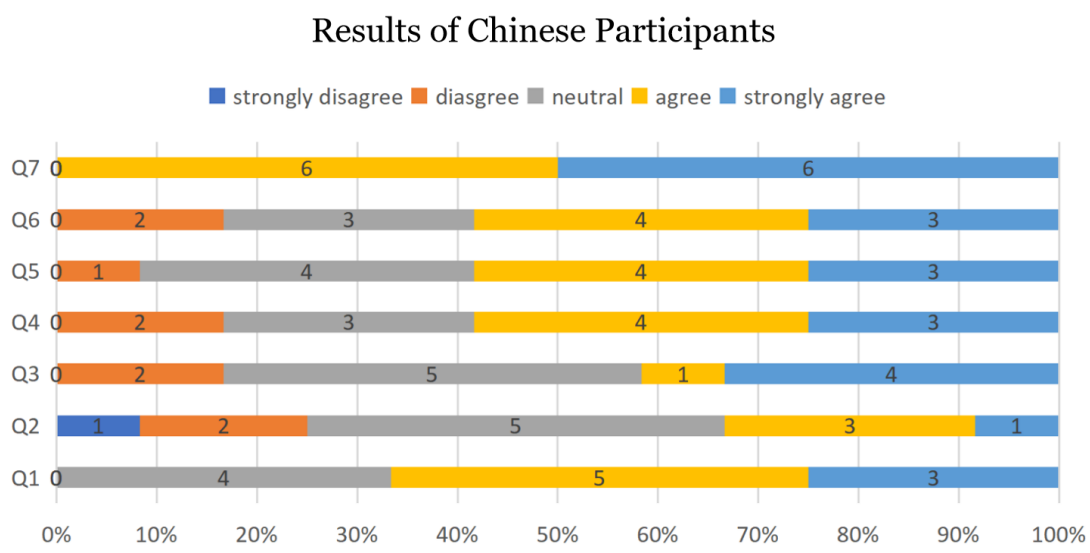


Figure 21: The distribution of scores assigned by Chinese participants across the seven questions.

References

[1] Stability-AI: stablediffusion. https://github.com/Stability-AI/StableDiffusion (2022)
[2] Hu, E.J., shen, Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. International Conference on Learning Representations (ICLR). https://openreview.net/forum?id=nZeVKeeFYf9 (2022)
[3] LingDong-: skeleton-tracing. https://github.com/LingDong-/skeleton-tracing (2020)
[4] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612. https://doi.org/10.1109/TIP.2003.819861 (2004)