2023 16th International Symposium on Computational Intelligence and Design (ISCID) | 979-8-3503-1745-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISCID59865.2023.00028

WordDiffuser: Helping Beginners Cognize and Memorize Glyph-Semantic Pairs Using Diffusion

Yuhang Xu[a],Wanxu Xia[b],Yipeng Chen[a],Jiageng Liu[a],Weitao You[a]*
[a]College of Computer Science and Technology, Zhejiang University , Hangzhou,PR China
[b] School of Cyber Science and Technology,Beihang University, Beijing,PR China
Email:Yuhang Xu@3200105216@zju.edu.cn, Weitao You@weitao_you@zju.edu.cn

***Abstract*—This paper delves into symbolic cognition, emphasizing the evolution from ancient hieroglyphs to modern abstract characters. Acknowledging the cultural importance and learning challenges associated with these abstract characters, we propose WordDiffuser, an AI-aided framework generating images to help beginners understand the relationship between abstract characters and their meanings. Our research includes a review of the role of semantic association in education, an investigation identifying market shortcomings, and the creation of an interactive learning website. The effectiveness of our approach is substantiated through an experiment involving 22 participants, showing considerable improvements in abstract character recognition and memorization. This study paves the way for further advancements in technology-assisted learning of symbolic systems.**[1]

***Index Terms*—AI-Generated Content, Symbolic, Stable Diffusion, Recognition, Characters**

I. Introduction

Symbolic cognition plays an important part in human development. A symbol refers to a specific image, and there are intricate links between symbols and semantics. The general definition of a symbol is that it exists independently and is linked to another, which can be “explained”.

The history of humans is accompanied by the history of symbols. Most of the primitive hieroglyphs are drawing symbols, which are expressed through painting. In the ancient Egyptian hieroglyphs, there were also several pictures to express the correspondence of things after the sound combination. Sumel’s hieroglyphs have also experienced this period. While in China, the earliest hieroglyph found was about 3000 years ago which was called Oracle. After thousands of years of history and evolution, these original Oracles still have a very close connection with the current Chinese characters.

Thus, the similarity between Oracles and current characters tell us the potential cultural bond between them, and we can find the significance of learning Oracles. Firstly, Research shows that the original Oracles always refer to observed things. As human thoughts continue to be enriched, complex and abstract vocabulary and grammar appear. Back to a simpler form, the cognitive process of original Oracles can reflect the paradigm of recognition better. Also, Oracles assist the recognition of current characters, and even other symbols. However, it proved difficult to recognize and memorize these Oracles and current characters for beginners, especially for young children. According to Saussure’s definition, Oracles(signifier) and items(signified) are glyph-semantic pairs. To memorize the glyph-semantic pairs clearly, high redundancy of logic connections is important.

The purpose of this paper is to find out the paradigm of character recognition for beginners and to establish logic connection in glyph-semantic pairs. The organization structure of the article is as follows. In part 2, We conducted a survey of related work including the application of glyph-semantic association in education, the integration of AI with education and presenting glyph-semantic generation with AI. In part 3, on the basis of formative investigations, inspired by Word-As-Image, we design the framework WordDiffuser which consists of stable diffusion and GPT-3.5 to generate various meaningful images and help beginners to establish the connection in glyph-semantic pairs. We also develop a website for interaction and experiment. In part 4, in order to prove our framework is sufficient to recognition and memorization, we carry out experiments with 22 participants about Oracle memorization and understand, and the results are obvious. Then we show the results in part 5.Eventually we come to a conclusion in part 6. Our contributions are as follows:

- We developed WordDiffuser, a framework that establishes connections between the semantic and visual aspects of words. This framework utilizes GPT-3.5 to enhance prompt information and employs LDM to guide image generation. WordDiffuser facilitates more efficient and rapid learning for symbol-based language learners, offering them an engaging and cost-effective learning approach.
- We created an interactive website dedicated to symbol-based language learning and conducted user interviews and UEQ questionnaires to evaluate the user experience. The results indicated high user satisfaction with our website, underscoring our contribution in providing an exceptional learning experience.

*Corresponding author.

[1] Our Demo website is avaliable at https://jiagengliu02.github.io/Demo/

2473-3547/23/$31.00 ©2023 IEEE
DOI 10.1109/ISCID59865.2023.00028

- We conducted a series of user experiments and conducted detailed analysis of the data. These experiments provided tangible experimental foundations for understanding the relationship between visual form and semantic meaning in symbol-based language learning. The insights gained from these experiments serve as a practical basis for future improvements and advancements in this domain.

II. RELATEDWORK

A. The Application of Glyph-Semantic Association in Education

Associating glyphs with semantics is a common approach in teaching. A survey by Liu Haio revealed that in grassroots preschool education, 58.5% of educators place substantial emphasis on teaching the sound and meaning of Chinese characters, with only 3.8% not prioritizing it. This demonstrates that linking glyph shapes with their semantic meanings is a crucial part of basic education.

Research by Gao Yuan and others shows that using pictographic characters in preschool education can enhance children's expressive thinking, imagination, and creativity, which traditional literacy methods cannot achieve. By combining character and image, we can effectively handle preschoolers' characteristics, such as lack of attention and emotional volatility, further improving their literacy skills and aesthetic experiences.

Experiments have definitively demonstrated the improvements brought about by using pictographic characters in teaching. Huang Xuemei's work empirically validated that Oracle Bone Script aids preschoolers' learning of modern Chinese characters and serves as a significant bridge. Regardless of children's interest, enthusiasm, imagination, or performance, literacy teaching using Oracle Bone Script is superior to conventional teaching. For some specific characters, such as "to maintain", the correct rate of teaching using Oracle Bone Script can increase by as much as 27%.

However, to integrate semantics with glyph shapes in teaching requires expert designers for image drawing and professional teachers for guidance and instruction. This is a major reason why this teaching method has not been widely adopted. If it is possible to automatically generate videos that link glyph shapes with semantics, this problem could be well addressed.

B. The Integration of AI with Education

In related studies, the application of Artificial Intelligence (AI) in the field of education is garnering increasing attention.

Baidoo-Anu and Owusu Ansah[3] focused on the potential benefits of the generative AI tool, ChatGPT, in education. They mentioned benefits of ChatGPT include but are not limited to promoting personalized and interactive learning, generating prompts for formative assessment activities, providing continuous feedback to guide teaching and learning, and so forth.

In fact, we have long used computers to assist in our learning[11], such as students using calculators, spreadsheets, and tools like MATLAB and Mathematica for calculations, statistical analysis, and simulations. We write papers and conduct research by consulting Wikipedia and online resources (blogs, social media, and academic articles), with the aid of powerful search engines such as Google and Google Scholar. Therefore, the application of AI in education is an observable trend. Su and Yang[17] proposed a framework applying ChatGPT in education to provide timely feedback to preschool teachers, thereby improving the quality of education. However, there is limited research on the introduction of deep learning techniques in an educational context; traditional AI technologies (like natural language processing) are widely adopted in educational settings, while more advanced techniques are seldom used[5]. Thus, our work, which utilizes AI to automatically generate videos linking glyph shapes with semantics, bears significance and a degree of innovation.

C. Presenting Glyph Semantics with AI

Most relevant to our goal are those works performing text semantic stylization. Tendulkar et al.[18] replace the characters in a given word with clipart icons describing a given topic. An autoencoder is used to measure the distance between the character and the desired category icon to select the most appropriate replacement icon. Similarly, Zhang et al.[21] replace similar stroke parts of one or more characters with clipart instances to generate decorative stylization. These methods operate in the raster domain and replace characters with existing icons, which limits them to predefined category sets in their dataset. However, our method operates in the vector domain and leverages the expressiveness of large pre-trained image-language models to create new illustrations that convey the desired concepts.

With the recent advancements in language-visual models[13] and diffusion models[12][14][15], the field of image generation and editing has experienced unprecedented evolution. These models, trained on millions of image-text pairs, have proven to effectively complete challenging visual-related tasks such as image segmentation[1], domain adaptation[16], image editing[2][7][20], personalization[6], and interpretability[4]. Despite being trained on raster images, their strong visual and semantic priors have also been proven to be successfully applicable to other domains such as motion[19], grids[10], and vector graphics. In our work, we similarly leverage the strong visual and semantic priors brought by the pre-trained steady diffusion model[15]. However, we maintain the relevance of font style and semantics based on this, thereby presenting glyph semantics with AI.

III. IMPLEMENTATION

A. Framework design

Framework is presented in Figure 2. User inputs a character in Chinese and our database in tff document will find the corresponding Oracle svg document S. We duplicate S as $\hat{S}$ to be the output that would be modified in each iteration. We apply Delaunay triangulation on them and calculate ACAPloss[8] between the two figure. Obviously, the original ACAPloss are

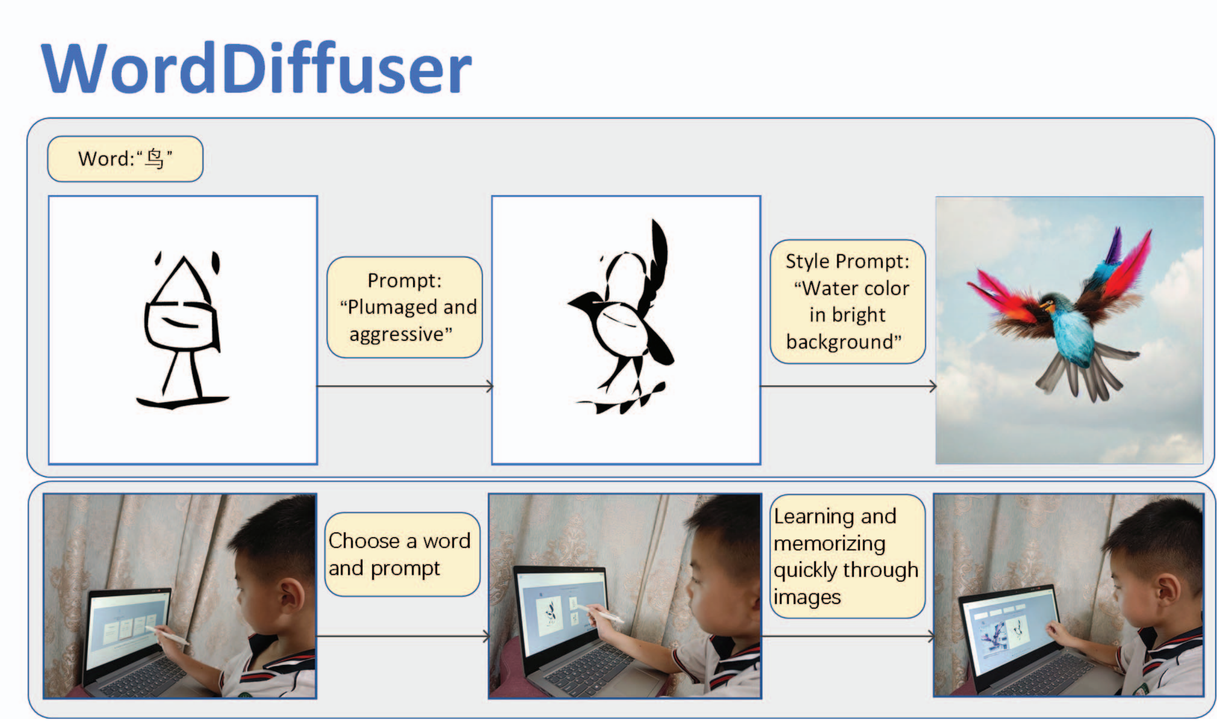

WordDiffuser
Word:"鸟"
Prompt: "Plumaged and aggressive"
Style Prompt: "Water color in bright background"
Choose a word and prompt
Learning and memorizing quickly through images

Fig. 1. Overview of WordDiffuser design and implementation. User input a word in Chinese, framework will retrieve the corresponding Oracle, following the chosen prompt to generate stage1 image, and following the chosen style prompt to generate stage2 image.Picture in the bottom is a beginner using our website to learn Oracles.

0 because S and $\hat{S}$ are the same. Then input both S and $\hat{S}$ into diffvg to get images P and $\hat{P}$ in pixel, inputting them into LPF to blur and calculate the TONEloss[8]. It represent the structure differences between two images and it also starts with 0. At the same time, pixel image $\hat{P}$ will be clipped or applied to other augment methods to get localized augmented image P_{aug}. In order to use LDM, P_{aug} pass through a pre-trained encoder and get the image Z in latent space. Then we add gaussian noise and input it into Unet to denoise, along with prompt for guiding image generation that are augmented by GPT-3.5. Finally, we get Generated image $\hat{Z}$ and calculate the difference with Z in latent space to get SDSloss. Eventually we combine the ACAPloss, the Toneloss and the SDSloss with configrated weight, using Adam optimizer to update $\hat{S}$. One iteration in stage1 has finished and after 500 iterations, $\hat{S}$ has been given a lot of semantic information, we input image in pixel $\hat{P}$ into stage1 output image I, so as to do the stylized generation of I. Stage2 simply apply image-to-image task of stable diffusion. Input I and style prompt will be set into two pre-trained specific encoder and fused in stable diffusion image-to-image API to generate the output image. input image I will be replaced by output image and then start the next iteration. we set iteration number of stage2 as 20 to get completely stylized images.

B. Dataset

we establish our dataset in font data part in our code. We use a spider to extract the tff files from websites and collect them into a dataset. Most common Chinese characters have correspond oracle characters in those tff files.

C. Website and UI Design

The front-end website built in this experiment has been released and developed based on HTML, CSS and JavaScript.The site design is based on the design concept referred in formative investigation:

- (1) High Attention to Symbolic Cognitive Education
- (2) Seeking Innovative Teaching Methods
- (3) Expectation for Technological Applications Rather Than Manual Design

Web UI design is shown in the figure below. Interaction is mainly divided into three stages:

- (1) Select an Oracle to learn, and select one of the corresponding four prompt after augmentation with GPT-3.5
- (2) View the first stage images generated by Oracle and prompt. In addition to the gif images, we also provide the first and last images of the gif changes. gif images are synthesized in 500 iterations
- (3) Select the corresponding image style according to personal preference, and view the style enhanced images generated in the second stage. We show the results of

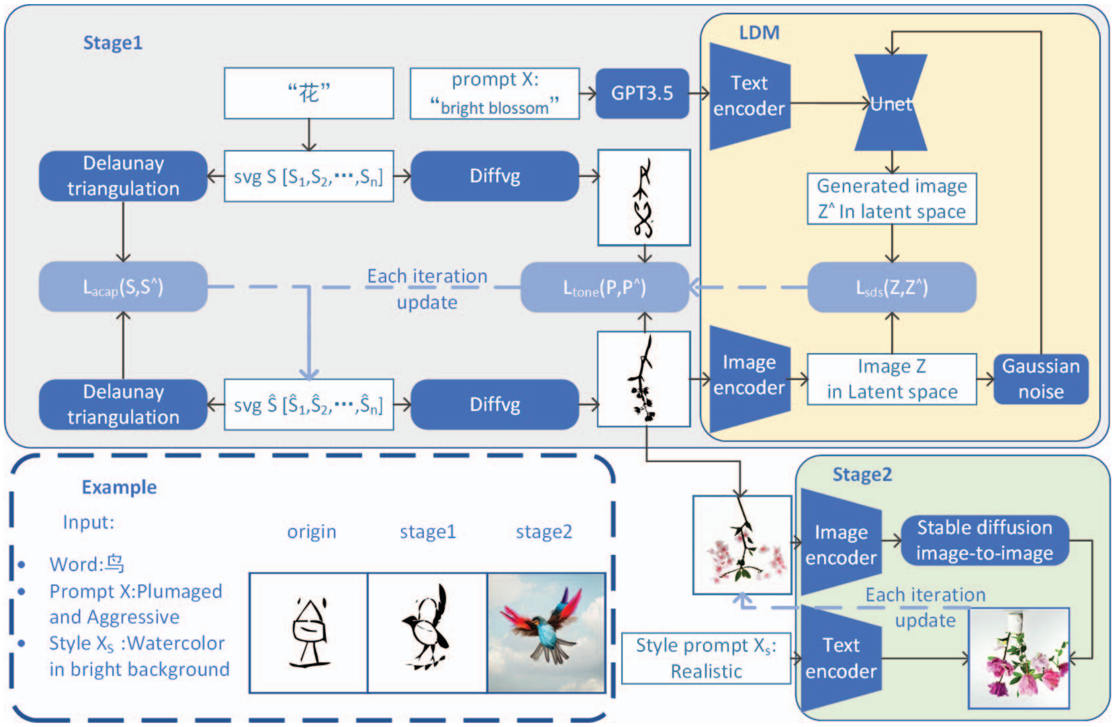
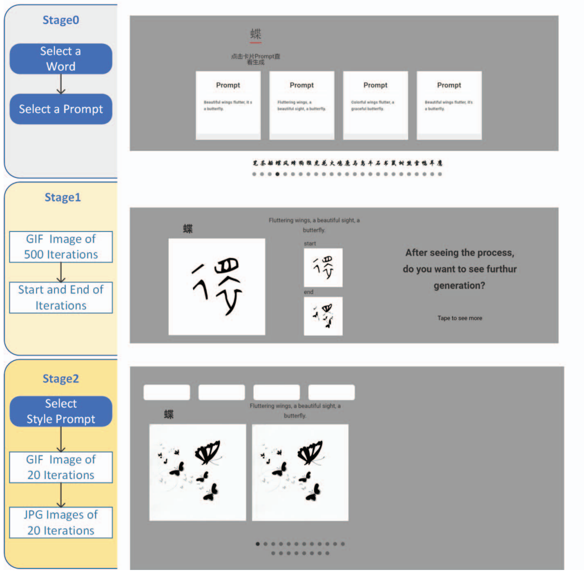


Fig. 2. framework design of WordDiffuser. The Chinese input will be converted into svg $S$and get a copy svg $\hat{S}$as output. In stage1, Firstly we use Delaunay triangulation to calculate the L_{acap} loss. Secondly we us Diffvg[9] to transform svg into image and calculate the L_{tone} loss. Thirdly we use LDM and GPT-3.5 augmented prompt to calculate the L_{sds} loss. The three loss are weighted summed and backpropagated to optimize svg $\hat{S}$. In stage2, we use Stable diffusion image to image pipeline to stylize and augment the output image of stage1.



Beautiful wings flutter, it s a butterfly.

Fluttering wings, a beautiful sight, a butterfly.

Colorful wings flutter, a graceful butterfly.

Beautiful wings flutter, it's a butterfly.

Fluttering wings, a beautiful sight, a butterfly.

After seeing the process, do you want to see furthur generation?

Fluttering wings, a beautiful sight, a butterfly.

Fig. 3. Website design

20 iterations of the respective image and the composite image

The whole process is unartifically accomplished by our framework, and the usage of 2 kind Stable Diffusion and GPT-3.5 corresponds to innovation. During (3) users are required to select different styles to meet their individual needs. To attract user's attention, the Web page has a lot of buttons to interact with and select different content.

IV. Experiment

In order to prove that the product we generated can indeed help users better learn the symbolic language, we designed the following experiments: We will divide the participants into three groups. The first group will learn through simple characters, the second group will learn through GIFs that change from glyphs to meanings, and the third group will learn based on the second group by changing from glyphs to meanings to objects in real life. After learning for the same period of time, we judge the learning effect by testing the memory accuracy of the newly learned characters of the subjects. In addition, to ensure that most people have no basic knowledge of the symbolic language we will be learning, we chose Oracle as the learning target for this experiment.

A. Paticipant

In this experiment, we recruited a total of 22 subjects. They were all college students between the ages of 19 and 22, equally male and female, and randomly divided into three groups. The participants were recruited through announcements from their personal social media accounts, ensuring that none of them had a background in Oracle learning and were native Chinese speakers with some experience in the symbolic language.

B. Experimental Process

1) Preparation: Because the experimenter is relatively scattered, we chose to conduct the experiment online. Before the experiment, we sent learning files to all subjects in advance, and gave them enough time to download and decompress them. However, they would not open them in advance for learning until we gave the instruction to start.

2) During the Experiment: When the experiment began, all subjects were asked to start the study at the same time. The learning content is a total of 20 oracle bones, and the learning files of each group, apart from the experimental content, we try our best to make no other differences. The study time was set at 5 minutes, which was the appropriate study length we got after the test. After the subjects finished learning, we asked all the subjects to stop learning at the same time, and then took a rest period of five minutes. The reason for not starting the accuracy test immediately is to ignore the effect of short-term memory, which can be retained for very short periods of time and does not really help learning. After the experiment, we let all the subjects try to experience our website freely, and asked them to fill in the UEQ questionnaire. We want to demonstrate that our product is fun and easy to use.

3) After the Experiment: On the second day after the experiment, we sent out the accuracy test questionnaire to the subjects again. This is to see if our product can work better over a slightly longer period of time. At the same time, we also selected several subjects to conduct post-experiment interviews, in order to have a detailed understanding of users' experience and feelings on this product.

V. Result

A. Result Confidence Analysis

TABLE I
Normality Test

	Kolmogorov-Smimov		**Shapiro-Wilk**	
	D	***p***	***W***	***p***
group	0.134	0.077	0.916	0.007
score	0.130	0.094	0.966	0.290

1) Normality Test: The SPSSAU website was used in this experiment, which is a mature and perfect statistical tool for processing and analyzing data. We first conducted normality test on experimental data (Table.1), which was to determine which method we would adopt in the subsequent data analysis.

TABLE II
Homogeneity Test of Variance

	group1	g1(n)	g2	g2(n)	g3	g3(n)	F	p
score	9.45	8.37	8.63	8.86	6.9	10.25	0.277	0.922

2) Homogeneity Test of Variance: As can be seen from the results of the homogeneity test of variance (Table.2), the data in the table showed no significance (p>0.05), so the homogeneity analysis of variance was satisfied. Normal distribution and homogeneity of variance are satisfied at the same time, so the final analysis method adopts the analysis of variance (ANOVA) in parameter test.

TABLE III
Analysis of Variance

	group1	g1(n)	g2	g2(n)
score	96.43±9.45	98.00±8.37	99.38±8.63	100.71±8.86
	g3	g3(n)	F	p
score	86.43±6.90	86.00±10.25	3.568	0.011

3) Analysis of Variance: According to the results of variance analysis (Table.3), it can be judged that there is a significant relationship between the groups and the scores (p<0.05), so it can be considered that the scores obtained by the experimenters in different groups are significantly different.

B. UEQ Analysis

As for the UEQ questionnaire, we analyzed it with the help of the data processing tool downloaded from its official website. The results show that our product is excellent in the degree of fun, but the subjects think it is still lacking in practicality, which may be caused by the fact that our results are mostly presented with pictures, which are too fancy. But overall, it's a good experience for users.

C. User Interview

After the experiment, we conducted a face-to-face interview with users and got the following feedback:

1) Significant Learning Gains: Users say that with the help of our products, it is really helpful to memorize oracle bone inscriptions quickly. Through visual capture of images to assist memory, familiar things become "strange" oracle bone inscriptions, so that memory can realize series interaction, which can improve the efficiency of oracle bone inscriptions memory to a certain extent. For oracle bone scripts with special fonts, some help will be provided. The fonts of these oracle bones are often quite different from the actual meaning, but through the final image of the product, it will be easy to leave a deep impression in a short time. "But also expressed some concerns about the product." The combination with Oracle is very attractive, but not sure how it will work with simplified Chinese characters.

2) Evaluation From a User Perspective: We asked respondents to judge the product from different perspectives. As a child learner, the user said, " Children are more stimulated by visual images. Novel, dynamic and intuitive things attract my attention more and make me more interested in understanding and learning." Parents, as guardians, feel they want their children to learn in this way, because "in my opinion, for children, 'interest' is the best teacher. This product can greatly promote children's interest in learning Chinese characters. If children can watch an image gradually change into Chinese characters like watching cartoons, they can gain both fun and knowledge during this process." And from the perspective of educators, they are also looking forward to using this product for teaching.

3) Shortcomings Experienced: At the same time, interviewees also put forward some of their own experience of the shortcomings of the part of the word generated by the image seems not vivid, or difficult to recognize at first sight, such as" snow "; Some words generate images that are a little off, like a horse with five legs. In the process of use, We think that auxiliary words can be used to help introduce the logical points of image changes. For example, it can be said that the components of a "boat" are "double OARS" and "sail". This introduction can further aid memorize. These are things we will improve in the future.

VI. CONCLUSION

Symbolic writing has a long history in human development, and there are a large number of users today. Just because of this, the study of symbolic characters has a stable demand today and in the foreseeable future. At the same time, one of the major features of characters is that its font and semantic are always closely related. If it is only memorized as a symbol text, it will inevitably affect the efficiency of learning and miss its cultural heritage. To this end, we developed a framework named WordDiffuser that can automatically generate the character pictures, adding more semantic features on the original fonts. After the experiment, we find that it does have a full gain effect on the learning and memory of symbols and characters. We believe that this product can link AI with symbolic education and help the new generation of learners to learn symbolic language more quickly and with more fun in the near future.

VII. ACKNOWLEDGEMENTS

This work is supported by the National Key R&D Program of China (2022YFB3303301), and the National Natural Science Foundation of China (No. 62107035).

REFERENCES

[1] Tomer Amit, Tal Shaharbany, Eliya Nachmani, and Lior Wolf. 2021. Segdiff: image segmentation with diffusion probabilistic models. *arXiv preprint arXiv:2112.00390*.

[2] Omri Avrahami, Dani Lischinski, and Ohad Fried. 2022. Blended diffusion for text-driven editing of natural images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18208–18218.

[3] David Baidoo-Anu and Leticia Owusu Ansah. 2023. Education in the era of generative artificial intelligence (ai): understanding the potential benefits of chatgpt in promoting teaching and learning. *Available at SSRN 4337484*.

[4] Hila Chefer, Shir Gur, and Lior Wolf. 2021. Transformer interpretability beyond attention visualization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 782–791.

[5] Xieling Chen, Haoran Xie, Di Zou, and Gwo-Jen Hwang. 2020. Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100002.

[6] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. An image is worth one word: personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*.

[7] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2022. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*.

[8] Shir Iluz, Yael Vinker, Amir Hertz, Daniel Berio, Daniel Cohen-Or, and Ariel Shamir. 2023. Word-as-image for semantic typography. *ACM Trans. Graph.*, 42, 4, Article 151, (July 2023), 11 pages. DOI: 10.1145/3592123.

[9] Tzu-Mao Li, Michal Lukáč, Gharbi Michaël, and Jonathan Ragan-Kelley. 2020. Differentiable vector graphics rasterization for editing and learning. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 39, 6, 193:1–193:15.

[10] Oscar Michel, Roi Bar-On, Richard Liu, Sagie Benaim, and Rana Hanocka. 2022. Text2mesh: text-driven neural stylization for meshes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13492–13502.

[11] San Murugesan and Aswani Kumar Cherukuri. 2023. The rise of generative artificial intelligence and its impact on education: the promises and perils. *Computer*, 56, 5, 116–121.

[12] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2021. Glide: towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*.

[13] Alec Radford et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[14] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*.

[15] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695.

[16] Kunpeng Song, Ligong Han, Bingchen Liu, Dimitris Metaxas, and Ahmed Elgammal. 2022. Diffusion guided domain adaptation of image generators. *arXiv preprint arXiv:2212.04473*.

[17] Jiahong Su and Weipeng Yang. 2023. Unlocking the power of chatgpt: a framework for applying generative ai in education. *ECNU Review of Education*, 20965311231168423.

[18] Purva Tendulkar, Kalpesh Krishna, Ramprasaath R Selvaraju, and Devi Parikh. 2019. Trick or treat: thematic reinforcement for artistic typography. *arXiv preprint arXiv:1903.07820*.

[19] Guy Tevet, Brian Gordon, Amir Hertz, Amit H Bermano, and Daniel Cohen-Or. 2022. Motionclip: exposing human motion generation to clip space. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII*. Springer, 358–374.

[20] Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. 2022. Plug-and-play diffusion features for text-driven image-to-image translation. *arXiv preprint arXiv:2211.12572*.

[21] Junsong Zhang, Yu Wang, Weiyi Xiao, and Zhenshan Luo. 2017. Synthesizing ornamental typefaces. In *Computer Graphics Forum* number 1. Vol. 36. Wiley Online Library, 64–75.