

# Utilizing stable diffusion and fine-tuning models in advertising production and logo creation: An application of text-to-image technology

**Chenyang Wang**

Jeme Tienyow Honors College, Beijing Jiaotong University, Beijing, 100091, China

20221020@bjtu.edu.cn

**Abstract.** This article delves into the implementation of text-to-image technology, taking advantage of stable diffusion and fine-tuning models, in the realms of advertising production and logo design. The conventional methods of production often encounter difficulties concerning cost, time constraints, and the task of locating suitable imagery. The solution suggested herein offers a more efficient and cost-effective alternative, enabling the generation of superior images and logos. The applied methodology is built around stable diffusion techniques, which employ variational autoencoders alongside diffusion models, yielding images based on textual prompts. In addition, the process is further refined by the application of fine-tuning models and adaptation processes using a Low-Rank Adaptation approach, which enhances the image generation procedure significantly. The Stable Diffusion Web User Interface offers an intuitive platform for users to navigate through various modes and settings. This strategy not only simplifies the production processes, but also decreases resource requirements, while providing ample flexibility and versatility in terms of image and logo creation. Results clearly illustrate the efficacy of the technique in producing appealing advertisements and logos. However, it is important to note some practical considerations, such as the quality of the final output and limitations inherent in text generation. Despite these potential hurdles, the use of artificial intelligence-generated content presents vast potential for transforming the advertising sector and digital content creation as a whole.

**Keywords:** stable diffusion, text-to-image technology, advertisement production, latent diffusion models.

## 1. Introduction

As the time-honored adage professes, "a picture is worth a thousand words." Indeed, images have an essential role in our daily lives. Ranging from the small scale of a book cover to the large scale of videos we consume, many elements of life comprise combinations of images. Particularly in the field of advertising, images possess a noteworthy capacity to seize people's attention far more effectively than words [1]. They consistently exhibit vibrant colors and compelling compositions, which command individuals' focus [2]. Moreover, a well-chosen image can directly communicate the features, functions, and advantages of a product or brand to the audience. The usage of symbols, icons, and visual elements can transmit messages, provoke emotions, and stimulate the desire to purchase [3].

Images possess the capacity to communicate information directly through visual means, thereby not necessitating the reliance on text [4]. Nonetheless, while there are considerable advantages to using images in advertising, creating an image is not always a convenient or cost-effective process. If you aim to incorporate models into your images to enhance their appeal, this entails significant expenses, such as initial conceptualization, preparation of photography equipment, and hiring of professionals like models, photographers, and lighting technicians. Alternatively, sourcing images from stock photo websites can be a time-intensive process and may not always yield suitable results. Even if one opts for software-generated images over real-world photographs, a professional designer is often still indispensable.

Thus, this article proposes an easier solution: generating suitable images through text-to-image technology and fine-tuning models. The images produced by this method can be effectively utilized in advertisements. Moreover, these fine-tuned models can be employed to create logos or generate creative ideas and inspiration. This approach can drastically reduce the need for human resources and material costs while producing images that align more closely with expectations.

## 2. Methodology

### 2.1. Stable diffusion technique

Stable diffusion is a technique widely used in the field of text-to-image and image-to-image recently. It is basically based on the Latent Diffusion Models, which are probabilistic generative models that achieve impressive synthesis results on high-resolution and complex scenes by decomposing the image formation process into a sequential application of denoising autoencoders [5].

It can be divided into two basic stages. The first stage is the compression learning stage, it uses variational autoencoders (VAE) to encode the input into a lower dimension. Through the compression learning stage, the model learns a low-dimensional latent representation space that is perceptually equivalent to the image space. This allows the subsequent generative learning stage to efficiently generate images in this low-dimensional space, without the need for complex computations in the high-dimensional image space [6]. In another word, the model pre-trained in this stage allows the users can do the text-to-image mission based on it.

The second stage is the generative learning. In this stage, the model utilizes the pre-trained low-dimensional latent representation space from the compression learning stage to generate images. This stage takes advantage of diffusion models (DMs), which are built from a hierarchy of denoising autoencoders. And the model architecture is enhanced with cross-attention layers, which enable the conditioning of the image generation process with inputs such as text or bounding boxes [7]. This means that users can provide additional information to guide the image synthesis, making it more flexible and versatile.

### 2.2. Fine-tuning and adaptation using LoRA

Low-Rank Adaption (LoRA) model was firstly proposed as a natural language processing model, it was used in the pre-trained large language models for fine-tuning in order to lead it into some specific tasks [8]. It allows the injection of trainable rank decomposition matrices into each layer of the Transformer architecture, effectively reducing the number of trainable parameters for downstream tasks.

The key principle behind LoRA is to freeze the pre-trained model weights and optimize the rank decomposition matrices instead. By doing so, LoRA achieves a significant reduction in the number of trainable parameters, making it more computationally efficient and memory-friendly. This addresses the challenge of deploying large language models, such as GPT-3, which have an enormous number of parameters and are prohibitively expensive to fine-tune independently, but now it is widely used in the process of the image generation [9].

Before the accomplishment of LoRA in the text-to-image, if the user wants to generate a specific element in an image, an anime character for example, it will be complicated and inaccurate.

However, the LoRA model offers a useful way to do such jobs, also with several advantages. Firstly, it allows for the sharing of a pre-trained model across multiple tasks. By freezing the shared model and only updating the rank decomposition matrices, the storage requirement and task-switching overhead are greatly reduced. Secondly, LoRA improves training efficiency by eliminating the need to calculate gradients or maintain optimizer states for most parameters. This lowers the hardware barrier to entry and enables faster training, especially when using adaptive optimizers [10].

Most Important, in the practical jobs it's twice as fast than Dreambooth method and sometimes even better performance than full fine-tuning, and the end result is also very small, which makes it easier to share and download.

The following four sets of images show the generated images of the target character, the generated images of the character without using LoRA, and the generated images of the character using LoRA under the same prompt. By comparison, it is evident that the use of LoRA produces better results.



**Figure 1.** The comparison of using Lora to generate images.

### 2.3. The Web-UI of stable diffusion

Stable Diffusion Web-UI can provide users a simple and powerful interface, which allows users to easily explore different modes and settings, and realize your creativity.

The users can utilize the text-to-image mode, allowing them to input any scene or object they can envision and have the model generate corresponding images for them. Many plugins, such as LoRA, can also be added to facilitate users in fine-tuning the generated images from various perspectives. Additionally, stable upscale mode can be used to enlarge and enhance low-resolution images, resulting in improved clarity and detail. In the following text, the Web-UI will be employed to ease the loading of various models, adjustments of prompt words, and utilization of plugins.

## 3. Approach

### 3.1. Streamlining production processes with stable diffusion and LoRA

**3.1.1. Advertisement.** The process of designing an outstanding advertisement graphic commences with the critical step of setting goals and identifying the intended audience. This initial phase lays the groundwork for the message the advertisement aims to convey and the audience it seeks to engage. Subsequently, market research is conducted to gain insights into the preferences, needs, and behavioral patterns of the target audience. This research also encompasses the examination of the advertising strategies and design styles employed by competitors. Such information is pivotal in better positioning the advertisement and crafting content that resonates with the audience.

Following the research, the creative conceptualization process begins. This involves brainstorming the theme of the advertisement, the method of information delivery, visual elements, and layout. The fundamental elements that are to be featured in the advertisement are identified and presented as prompts

in the Web-UI. The selection of suitable images and text is the next crucial step. The images, generated in bulk, are sifted through to find those that align with the advertisement's theme and appeal to the target audience. These images should be clear, attractive, and relevant. The diversity in style is achieved by adjusting different model parameters or prompt weights, based on the generated images. For content with distinct characteristics or styles, fine-tuning with the LoRA model can lead to optimal results.

Once a satisfactory image is obtained, it is further optimized and adjusted using tools such as Photoshop. This stage involves refining the details, and designing the font placement and layout to enhance the advertisement's visual appeal. The goal is to ensure the layout is clear, easily legible, and effectively guides the audience's attention.

Upon completion of the final design, the advertisement is exported in an appropriate file format, ready for publication across various media channels. Adjustments in size and format may be required to cater to the specific requirements of each advertising channel. This comprehensive process ensures the creation of an engaging and visually appealing advertisement graphic.

**3.1.2. Logo.** The process of designing a compelling logo commences with clearly defining the target client. Primarily, this includes understanding the brand - its products or services, target audience, and the key message they aim to convey through the logo. The inherent meaning that the logo needs to express is consequently outlined.

The next stage involves creative ideation, generating preliminary design elements and prompts corresponding to the brand's unique attributes. Here, the LoRA model is instrumental in forming the logo style based on these components. The primary elements are incorporated in the form of words or phrases separated by commas and form the basis of prompts added in the relevant areas of the Web-UI.

Subsequently, the crux of appropriate logo selection takes place. The process requires curating images from an array of generated options that cater to the client's needs. The image should be crisp, aesthetically pleasing, simplistic, and engaging. Additionally, diversification of style can be achieved by altering models or prompt weights based on the qualities of the generated images. The multitude of images can also provide inspiration, aiding in creating a more polished logo.

After obtaining a satisfactory logo, optimization and adjustments follow. Tools akin to Photoshop can be utilized for refining details and adding limited textual content to the logo to align it with the ideal version. This assures that the final product surpasses the initial concept, thereby producing an impactful and high-quality logo which conveys the brand's identity and values succinctly. This comprehensive approach to logo design ensures that the end product is not just visually appealing but encapsulates the essence of the brand successfully.

### **3.2. Aims and benefits of the approach**

In the abstract of generating images and logos for advertisements, using the StableDiffusion AI model has numerous aims and benefits, specifically when incorporating the text-to-image technique and fine-tuning of the model. This approach fundamentally aims to streamline the image creation process, minimizing the need for extensive resources and materially intensive methods.

The main objective of this approach is to offer a convenient, cost-effective, and time-efficient solution. The utilization of the Stable Diffusion AI model eradicates the need for expensive processes, such as conceptualization, preparatory steps for photography, and outsourcing professionals like models, photographers, and lighting technicians. It also reduces dependence on stock photo websites, which can be time-consuming and might not yield suitable images.

The benefits of this model implement a value-added system that ensures quality and adaptability. It can generate images that closely align with the user's expectations while simulating the capabilities of a professional designer, providing high-quality, versatile image outputs and logos that can be customized to suit varying requirements. The fine-tuned models can be used not just for generating creative images but also for developing logos and deriving inspirations.

Furthermore, it provides flexibility and opens a realm of possibilities for advertising agencies and designers alike. By just inputting prompts, they can receive an array of image and logo options. Each

option can offer a unique perspective that reflects the desired message or feeling. This not only fast-tracks the creative process but also allows for exploration of diverse options without the constraints of traditional design methods.

Thus, the application of the Stable Diffusion AI model for generating advertisement images and logos through the text-to-image technique and fine-tuning strategy can revolutionize the creative designing process. It simplifies, economizes, and optimizes image creation while guaranteeing delivery of compelling visuals that encapsulate the intended messaging effectively.

## 4. Results

### 4.1. Real-world applications

*4.1.1. Advertisement.* Considering beer is the world-wide popular drinking, it is used as an example of the advertisement in this case. In the Results section of this study, the practical application of the Stable Diffusion AI model in generating images for a beer advertisement is explored. The primary objective is to create an appealing image that not only represents the product - beer, but also illustrates its refreshing quality. Certain elements such as condensation on the beer bottle and beer foam are included in the image to enhance its visual appeal and make it more enticing to potential customers. Consequently, the keywords for the prompt have been determined.

The selected keywords are 'beer', 'condensation', and 'beer foam'. These are fed into the AI model as prompts for image generation. The goal is to produce an image that effectively showcases a cold, refreshing beer with foam. Moreover, there is another key prompt creatively chosen to be added, 'snow'. It makes the visual elements of the image more diverse and rich, thereby creating an enticing visual representation that would appeal to the target audience's senses, and ultimately, their desire to purchase and consume the product. As shown in Figure 1.



**Figure 2.** Advertisements (Photo/Picture credit: Original).

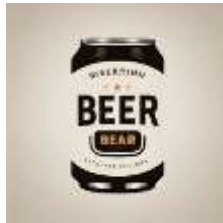
*4.1.2. Logo.* Retaining the context from the previous sections where beer was used as the example, the same approach is applied in the creation of a logo. For this purpose, it is paramount for the logo to exhibit simplicity and clarity. In pursuit of such stable output, the 'NeverEnding Dream' checkpoint is adopted. This large model, known for generating 3D-style imagery, ensures the elimination of excessive and unnecessary details, thereby giving rise to a smoother and more cohesive image.

Moreover, in order to produce images exhibiting distinct logo traits such as uncluttered backgrounds primarily in solid colors and a prominent subject boasting unique geometric features, the LoRA model 'Anylogo' is engaged. The prompts "logo" and "beer" have been utilized here. Examining the resultant images, the following can be observed. As shown in Figure 2.



**Figure 3.** Logos (Photo/Picture credit: Original).

The results clearly demonstrate that the combination of these two models brings forth aesthetically appealing logo images. However, the text appearing in most logos does not carry explicit meaning. This occurrence is attributed to the underlying technology's inability to recognize or generate words. It can only learn to produce shapes resembling letters. Therefore, post AI generation, further adjustments to the images are necessitated, generally via software like Photoshop. The process is illustrated using one of the generated images as an instance. As shown in Figure 3 and 4.



**Figure 4.** Original logo (Photo/Picture credit: Original).



**Figure 5.** Logo without wrong texts (Photo/Picture credit: Original).

#### *4.2. Implementation considerations*

The capabilities of Stable Diffusion's AIGC technology have been instrumental in generating commendable advertisements and logos. However, a rigorous in-depth analysis has uncovered several areas that require further attention and consideration to refine the process.

One such area is the quality of the output generated by AI models. While AI models can produce a wide variety of images, the inconsistency in quality often presents significant challenges. To maximize the effectiveness of these generated images, it is crucial to ensure that they meet the standards and requirements of specific use-cases. Therefore, the development of an objective evaluation scheme, potentially drawing upon principles from mathematics, is essential for comparing the generated content with predetermined quality standards.

Another important aspect to refine is the selection of prompts in the generation process itself. The model should be sensitive enough to choose precise prompts and adjust them promptly based on the output of each iteration. When executed meticulously, this can have a dramatic impact on the quality of

subsequent rounds. By closely examining the shortcomings of each iterative step, it becomes possible to fine-tune the succeeding prompts, thereby ensuring a continuously improving narrative.

Turning attention to text generation, it is evident that our AI models currently have limitations. As observed in the logo generation process, the ability to generate meaningful text is somewhat restricted and often requires additional editing and adjustments once the generation is complete. This additional step in the process highlights the need for improvement and should be approached with an initiative to further evolve the models. By recognizing and addressing these concerns, users can facilitate the maturation process of our AI models, ensuring they continue to produce outputs of exceptional quality.

## 5. Conclusion

In conclusion, while the realm of Artificial Intelligence Generated Content (AIGC) is experiencing rapid growth, it is still in its infancy, with many areas yet to be fully explored and developed. One of the chief concerns revolves around the lack of comprehensive legal regulations pertaining to AIGC. The copyright issues associated with AIGC content are ambiguously defined across various countries and regions, creating a potential breeding ground for legal disputes.

Moreover, the credibility of AIGC content raises additional concerns. The authenticity of AI-generated images cannot be assured, potentially undermining their effectiveness in certain applications. For example, in the fashion industry, substituting real models with AI might reduce costs, but the resultant images could arouse skepticism regarding their authenticity. This could inadvertently counteract the original intention of helping customers visualize the actual impact of a product. Nevertheless, despite these challenges, the incredible potential of this technology is undeniable. When harnessed correctly, AI models like Stable Diffusion used for generating images and logos can trigger a revolution in the advertising industry, drastically reducing costs and human resource requirements. Furthermore, it can offer a more efficient and flexible approach to creating visually captivating content that aligns closely with a brand's identity and message. As this technology continues to evolve, it is anticipated that the issues surrounding its application will be addressed, and more robust legal frameworks will be established. Without a doubt, with cautious and ethical usage, AIGC technology is poised to generate substantial economic benefits in the future, revolutionizing various industries. This makes it a promising instrument in the arena of digital content creation.

## References

- [1] Vahid, H., & Esmae'li, S. (2012). The power behind images: Advertisement discourse in focus. *International Journal of Linguistics*, 4(4), 36-51.
- [2] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10 684-10 695).
- [3] Hu, E. J., Shen, Y., Wallis, P., et al. (2021). Lora: Low-rank adaptation of large language models. *arXiv preprint. arXiv:2106.09685*.
- [4] Valipour, M., Rezagholizadeh, M., Kobzyev, I., & Ghodsi, A. (2022). Dylora: Parameter efficient tuning of pre-trained models using dynamic search-free low-rank adaptation. *arXiv preprint arXiv:2210.07558*.
- [5] Rost, M., & Andreasson, S. (2023). Stable Walk: An interactive environment for exploring Stable Diffusion outputs.
- [6] Pieters, R., Wedel, M., & Zhang, J. (2007). Optimal feature advertising design under competitive clutter. *Management Science*, 53(11), 1815-1828.
- [7] Malik, M. E., Ghafoor, M. M., Iqbal, H. K., et al. (2013). Impact of brand image and advertisement on consumer buying behavior. *World Applied Sciences Journal*, 23(1), 117-122.
- [8] Wu, J., Gan, W., Chen, Z., et al. (2023). AI-generated content (aigc): A survey. *arXiv preprint. arXiv:2304.06632*.
- [9] Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*, 90.

- [10] Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). A survey of Generative AI Applications. arXiv preprint. arXiv:2306.02781.