

# Application of Semantic Image Generation Techniques Based Detail Preserving Image Method for Repoussé Craft

Hong-En Shao\*

*School of Fashion Design, Shanghai Institution of Visual Arts, Wen Xiang Road,  
Shanghai 201620, China*

---

## Abstract

Text-to-image technology is a technology for mapping target image sets through natural language and is the latest production method for new content. The repoussé craft<sup>1</sup> is ancient Chinese gold and silver fine art crafts. Although the craft has been studied, its exploration in digital research is relatively new. To modernise the repoussé craft, this paper takes the combination of 3D generated drawings of repoussé craft motifs with text-to-image technology as a research object, applies current artificial intelligence to the study and analyses the possibilities of the combined application and related text specification recommendations, tests and analyses the impact of the changes in the prompt and weighted value settings on the repoussé craft jewelry design style. The results of the study show that the way the prompt<sup>2</sup> and matting image weights are set plays a key role in the stabilisation of the imaging style. This study not only fills the gap in the combination of repoussé technology and AI technology but also plays a positive role in the stability of the style of the enhanced text to imaging, which provides a specific parameter basis and a new way of thinking for the future application of AI technology in the design of fashion jewelry and even in the wider field of fashion design.

*Keywords:* Artificial intelligence; Repoussé craft; Unity of style; Fashion jewelry design

---

## 1 Introduction

The repoussé craft is an ancient gold and silver joinery process in which fine and complex patterns are achieved through the process of hammering and burin striking on the front and back of a relatively smooth and flat metal to create texture and surface effects such as lines, reliefs or intaglio carvings [1]. Its motifs and ornaments are core for expressing national cultural characteristics

---

<sup>1</sup> Repoussé craft is a process of hammering a burin with a set of basic patterns to create a pattern of unevenness and convexity on a metal surface. Burin is an ancient craft.

<sup>2</sup> Prompt is a short textbook phrase that generates a picture.

\*Corresponding author.

*Email address:* 1094386933@qq.com (Hong-En Shao).

and aesthetic sensibilities. However, the craft is in decline due to its complexity, and the motifs are too traditional and lacking in innovation to appeal to a wider and younger consumer group [2]. Currently, the Chinese repoussé craft is in desperate need of innovation in design content to regain popularity and resonance with the public.

Today, in computer vision and natural language processing, the continuous development of deep learning techniques for image generation with convolutional neural networks has led to the development of many deep network models for semantic-based image generation [3]. Although Text-to-Image is quite recent, its results have been remarkable and have led to a new wave of research in this field. Semantic-based image generation technology maps natural language and image set features to generate corresponding images based on natural language descriptions. It also uses linguistic properties to achieve purposeful representation of visual images in a generic, flexible and intelligent way [4]. Currently, the most representative AI semantic image generation tools include Stable Diffusion 2.0<sup>3</sup>, DALL-E<sup>4</sup> and Mid Journey 5.2<sup>5</sup>. As shown in (Fig. 1 and Fig. 2), Mid Journey 5.2 was used to input the descriptor 'ruby silver bunny ring'. The resulting images show that the technique has a high degree of textual differentiation and generalisation of the target visual attribute descriptions, which results in excellent visual performance in accuracy, resolution, variety and viewability.

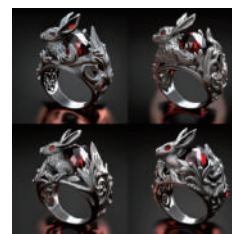
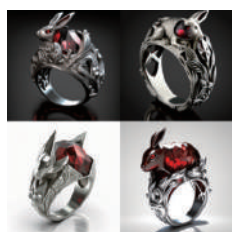


Fig. 1: Generation of Midjourney design scheme      Fig. 2: Optimization of Midjourney design scheme

Until 2021, the field of text-to-image generation is largely realised based on Generative Adversarial Networks GANs. For example, Mai Cong Hung, Osaka University, Japan, uses Generative Adversarial Networks (GANs) artificial intelligence techniques to analyse differences between artworks and art genres [5]. After 2021, more deep learning models that are independent of the GAN logic system have been proposed and have been well received, with results as good as, if not better than, those of GAN (Fig. 3). This model is capable of creating a new image from scratch while guaranteeing consistency with the text content, and can regenerate any rectangular region of an existing image [6]. For example, Yunfei Li of Qinghai University, in his master's thesis "Research on Sketch-Based Image Generation Techniques" in 2022, researched to achieve the task of drawing a thangka image with a simple sketch using ternary networks and the basic idea of generative adversarial networks [7]. As can be seen, semantic image generation is a collision of technology and art that has arrived.

Although text-to-image generation technology is in the process of iterative updating at a rapid

<sup>3</sup>Stable Diffusion is a latent text-to-image diffusion model capable of generating photo-realistic images given any text input, cultivates

autonomous freedom to produce incredible imagery, empowers billions of people to create stunning art within seconds.

<sup>4</sup>DALL-E 2 is an AI system that can create realistic images and art from a description in natural language.

<sup>5</sup>Midjourney is an AI painting platform which is an independent research lab exploring new mediums of thought and expanding the imaginative powers of the human species.