# SkipInject: Expanding Control of Diffusion Models Leveraging the Models Themselves

Ludovica Schaerf[1,*]

[1]*Digital Visual Studies, UZH-MPG, Zurich, Switzerland*

## Abstract

The landscape of computational design has evolved rapidly with the advent of generative models, such as diffusion models, which have revolutionized image creation and editing across diverse creative domains. These models, particularly Stable Diffusion [1] and its variants, provide designers with powerful tools that blur the line between creation and conception, opening new opportunities for creative experimentation. This paper explores an innovative approach to image editing that leverages the internal architecture of Stable Diffusion, specifically the skip connections of its U-Net [2] backbone, to enable training-free, flexible content and style transfer. This research contributes to the growing field of computational creativity, offering fresh perspectives on how computational methods can be employed to manipulate design artifacts in novel ways.

Computational design traditionally involves automation of design processes, the creation of custom tools, and the extension of stylistic forms. However, the integration of generative models into the design workflow moves beyond mere automation, positioning the computer as a collaborator in the conceptualization of creative works. While recent advances in diffusion models have largely focused on fine-tuning or retraining models for specific tasks, our approach utilizes the inherent flexibility of the U-Net architecture, particularly its skip connections, to offer a more controlled and efficient means of editing images. Unlike prior methods that rely on training models from scratch or fine-tuning existing models, our approach, SkipInject [3], enables precise content and style transfer by injecting the spatial information from one image into another, preserving the core structural elements while modifying stylistic features.

We systematically investigate the role of skip connections within the Stable Diffusion model, analyzing how these connections contribute to the separation of content and style in the image generation process. Through a series of experiments, we demonstrate that the third encoder-decoder block in the U-Net architecture plays a crucial role in disentangling content from style, making it possible to transfer these elements independently. Our method offers significant improvements over state-of-the-art techniques, achieving superior content alignment and structural preservation in image editing tasks.

The creative potential of this method extends beyond basic style transfer. By modulating the intensity of the content and style blending process, designers can fine-tune the extent of transformation, allowing for a more iterative, exploratory approach to design. Furthermore, we introduce three modulation techniques - classifier-free guidance, depth-wise alternation, and timestep manipulation - that provide additional layers of control, enhancing the flexibility of the editing process. These techniques make the method adaptable to a wide range of creative tasks, from subtle, detail-oriented modifications to more radical, transformative changes.

Our experiments cover a broad spectrum of image editing tasks, including text-guided image manipulation, style transfer, and fine-grained feature editing, demonstrating that our method achieves state-of-the-art performance on various benchmarks. Qualitative and quantitative evaluations show that SkipInject outperforms other leading methods in both text fidelity and structural preservation, offering a more robust and controllable approach to image editing.

The findings from this paper suggest that diffusion models, and particularly the U-Net architecture, can serve as a foundation for more intuitive, flexible, and creative design practices, with wide-ranging applications in fields such as visual arts, graphic design, and multimedia. In conclusion, this paper presents a novel, efficient, and controlled method for image editing and style transfer that leverages the skip connections of Stable Diffusion. Through systematic exploration and experimentation, we provide new insights into the inner workings of U-Net-based diffusion models, offering tools for creative professionals to explore and manipulate design artifacts in ways that were previously unattainable. As generative models continue to advance, the possibilities for computational creativity are vast, and this research serves as a step toward more dynamic and participatory design processes.

## Keywords

Diffusion Models, Skip Connections, Content-Style Disentanglement,

*WCDCC 2025: First Workshop on Computational Design and Computer-aided Creativity 2025, 23 June, 2025, Campinas, BR*

*Work accepted at CVPR 2025 CVEU Workshop.

ludovica.schaerf@uzh.ch (L. Schaerf)

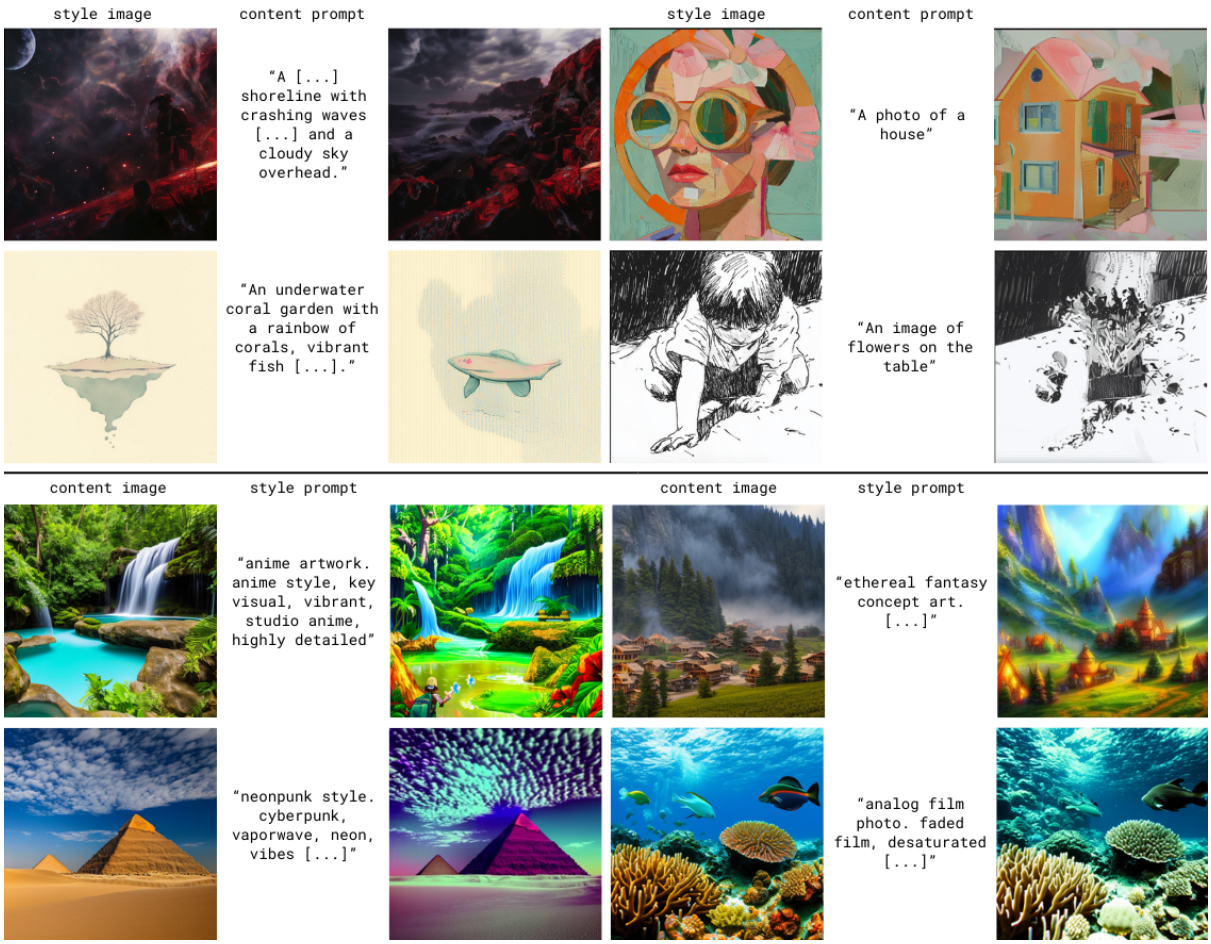


**Figure 1:** Example results from the original paper. The AI Art online communities offer an incredible wealth of information on style transfer in blogs such as Stable Diffusion Art that could be leveraged to build applied benchmarks for style transfer. In this figure, we show two interesting applications of our method: the first consists of the transfer of closed-source styles (e.g., styles used in Midjourney) to Stable Diffusion outputs using single-image style transfer (on the left). The second leverages the style prompts (with respective negative prompts) released by StabilityAI to transfer the described styles to real images or selected generated images (on the right).

## Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 in order to: Paraphrase and reword. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

## References

[1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10684–10695.

[2] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241.

[3] L. Schaerf, A. Alfarano, F. Silvestri, L. Impett, Training-free style and content transfer by leveraging u-net skip connections in stable diffusion 2., arXiv preprint arXiv:2501.14524 (2025).