VSF: Simple, Efficient, and Effective Negative Guidance in Few-Step Image Generation Models By Value Sign Flip

Wenqi Guo [*]**& Shan Du** [†]
Department of CMPS
University of British Columbia
Kelowna, BC V1V 1V8, Canada
wg25r@student.ubc.ca,shan.du@ubc.ca

Abstract

We introduce Value Sign Flip (VSF), a simple and efficient method for incorporating negative prompt guidance in few-step diffusion and flow-matching image generation models. Unlike existing approaches such as classifier-free guidance (CFG), NASA, and NAG, VSF dynamically suppresses undesired content by flipping the sign of attention values from negative prompts. Our method requires only a small computational overhead and integrates effectively with MMDiT-style architectures such as Stable Diffusion 3.5 Turbo, as well as cross-attention-based models like Wan. We validate VSF on challenging datasets with complex prompt pairs and demonstrate superior performance in both static image and video generation tasks. Experimental results show that VSF significantly improves negative prompt adherence compared to prior methods in few-step models, and even in CFG in non-few-step models, while maintaining competitive image quality. Code and ComfyUI node are available after publication.

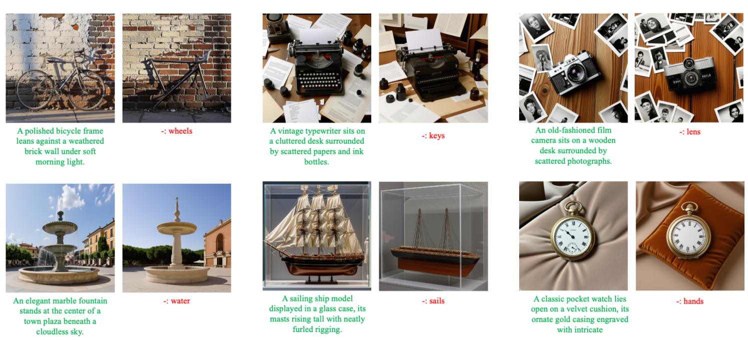

Figure 1: Original image without negative guidance and image generated using our VSF negative guidance on Stable Diffusion 3.5 Large Turbo. The green prompt is the positive prompt, and the red one is the negative prompt. These examples have significant changes as they are removing essential parts of an object.

[*]Also affilated to Weathon Software
[†]Corresponding author

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: The First Workshop on Generative and Protective AI for Content Creation.

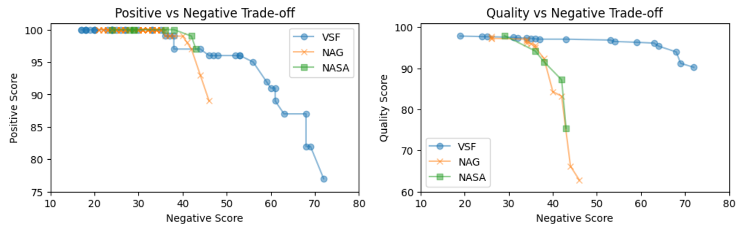

Figure 2: Trade off plot of positive-negative score and quality-negative score. Both axies follows "higher is better."

	Positive Score	Negative Score	Quality Score
VSF	0.870	0.545	0.952
NAG (Chen et al., 2025)	0.993	0.220 ($p < 10^{-6}$)	0.968
NAG++	0.975	0.320 ($p < 10^{-6}$)	0.901
NASA(Nguyen et al., 2024)	0.970	0.380 ($p < 10^{-6}$)	0.867
None	0.990	0.195 ($p < 10^{-6}$)	0.968
CFG (Ho & Salimans, 2022) (non-few-step)	1.000	0.300 ($p < 10^{-6}$)	0.956

Table 1: Positive scores (how well the model follows the positive prompts) and negative scores (how well the model avoids the negative prompts) of our model (VSF), NAG (Chen et al., 2025), and NAG with hyperparameter re-tuned (NAG++).

1 Results

1.1 Dataset

Following Park et al. (2025), we use ChatGPT o3 (Open AI) to generate pairs of prompts and negative prompts. Unlike prior work, our prompts are intentionally more challenging: the negative prompt is typically related to the positive one, and as a critical component—e.g., the positive prompt of a bike could have a negative prompt of "wheels". Additional examples are shown in Figure 3. Besides prompts, two questions are generated at the same time for later evaluation, one query if the image has the main object, either with or without the negative element, and the other one queries if the negative prompt element is missing. Prompts are generated in batches. Due to the fact that the model might output similar concepts for different prompts, it may introduce some repetition across batches or within batches with different phasing. There are 200 prompts generated, and we run them with 2 different seeds for the main results.

1.2 Baseline

We chose NAG (Chen et al., 2025) and NASA Nguyen et al. (2024) as our baseline. We also used a base model without negative guidance as a bare baseline, aiming to show the lower bound of the dataset (i.e., how likely the positive prompt will introduce the negative concept, if there is no negative guidance). Because NASA's original source code was not publicly available at the time of writing, we reimplemented it based on NAG's codebase. Specifically, we replaced the guidance equation from NAG with NASA's equation, removed normalization and blending, and enabled guidance when the scale is greater than 0 (instead of 1). Additionally, to compare our method in few-step models with CFG on original non-few-step models, we also used the original Stable Diffusion 3.5 Large with CFG as a baseline.

1.3 Metric

Following Park et al. (2025); Wei et al. (2025), we used multimodal large language models (MLLM), specifically `llama-4-maverick-17b-128e-instruct-fp8`, to evaluate if the generated image follows the positive prompt and the negative prompt using the two questions generated during prompt generation. LLaMA 4 Maverick has a very high image reasoning MMMU Yue et al. (2024) score, higher than Gemma 3 and even GPT-4o. We avoided using the same model (o3) for both evaluation and generation for cost control and to avoid bias within a model.

References

Dar-Yen Chen, Hmrishav Bandyopadhyay, Kai Zou, and Yi-Zhe Song. Normalized Attention Guidance: Universal Negative Guidance for Diffusion Models, June 2025. URL http://arxiv.org/abs/2505.21179. arXiv:2505.21179.

Jonathan Ho and Tim Salimans. Classifier-Free Diffusion Guidance, July 2022. URL http://arxiv.org/abs/2207.12598. arXiv:2207.12598.

Viet Nguyen, Anh Nguyen, Trung Dao, Khoi Nguyen, Cuong Pham, Toan Tran, and Anh Tran. SNOOPI: Supercharged One-step Diffusion Distillation with Proper Guidance, December 2024. URL http://arxiv.org/abs/2412.02687. arXiv:2412.02687.

Open AI. Introducing OpenAI o3 and o4-mini. URL https://openai.com/index/introducing-o3-and-o4-mini/.

Junsung Park, Jungbeom Lee, Jongyoon Song, Sangwon Yu, Dahuin Jung, and Sungroh Yoon. Know "No" Better: A Data-Driven Approach for Enhancing Negation Awareness in CLIP, March 2025. URL http://arxiv.org/abs/2501.10913. arXiv:2501.10913.

Xinyu Wei, Jinrui Zhang, Zeqing Wang, Hongyang Wei, Zhen Guo, and Lei Zhang. TIIF-Bench: How Does Your T2I Model Follow Your Instructions?, June 2025. URL http://arxiv.org/abs/2506.02161. arXiv:2506.02161.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, and Wenhu Chen. MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI, June 2024. URL http://arxiv.org/abs/2311.16502. arXiv:2311.16502.

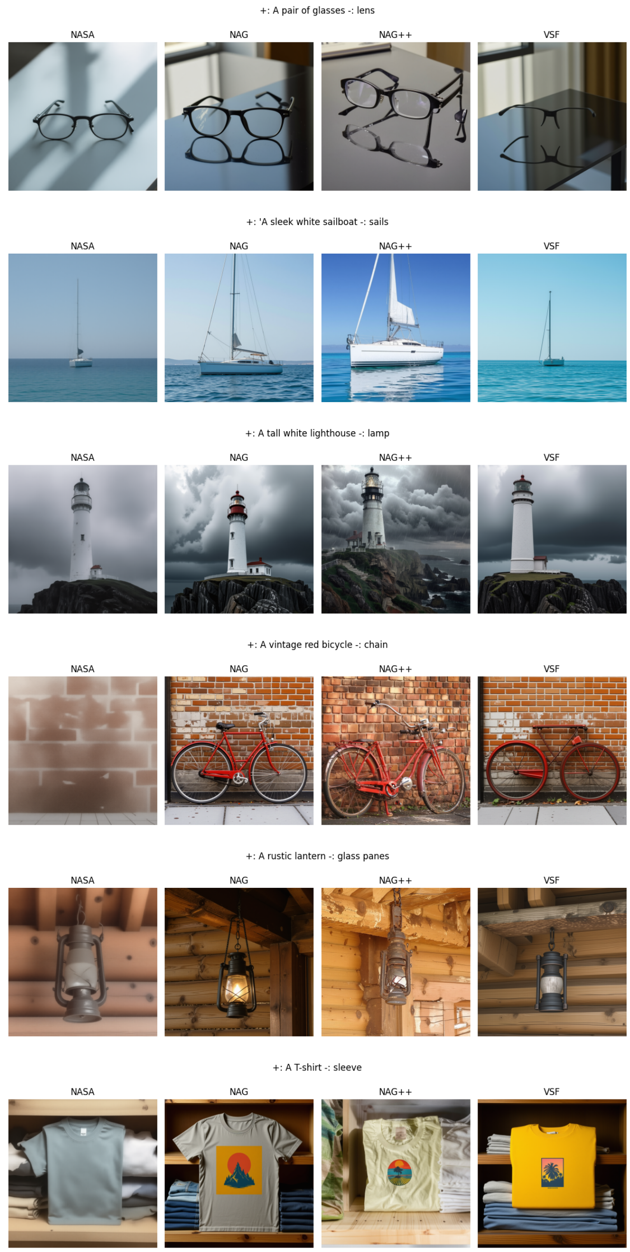

Figure 3: Selected Results for Comparison. Positive promtes are condensed for spacing.

A More Examples

In Figure 3 we showed a comparison between our method and previous methods. In Figure 4, we showed some examples of using our negative guidance to push away the image from a generalized "good" image, creating special art styles. In Figure 5 we showed some examples where we use the same noun in both positive and negative prompts, semi-canceling the item's appearance and thus creating very abstract art. Video examples are shown in `https://vsf.weasoft.com/`. More results are in our pre-print `https://arxiv.org/abs/2508.10931`.

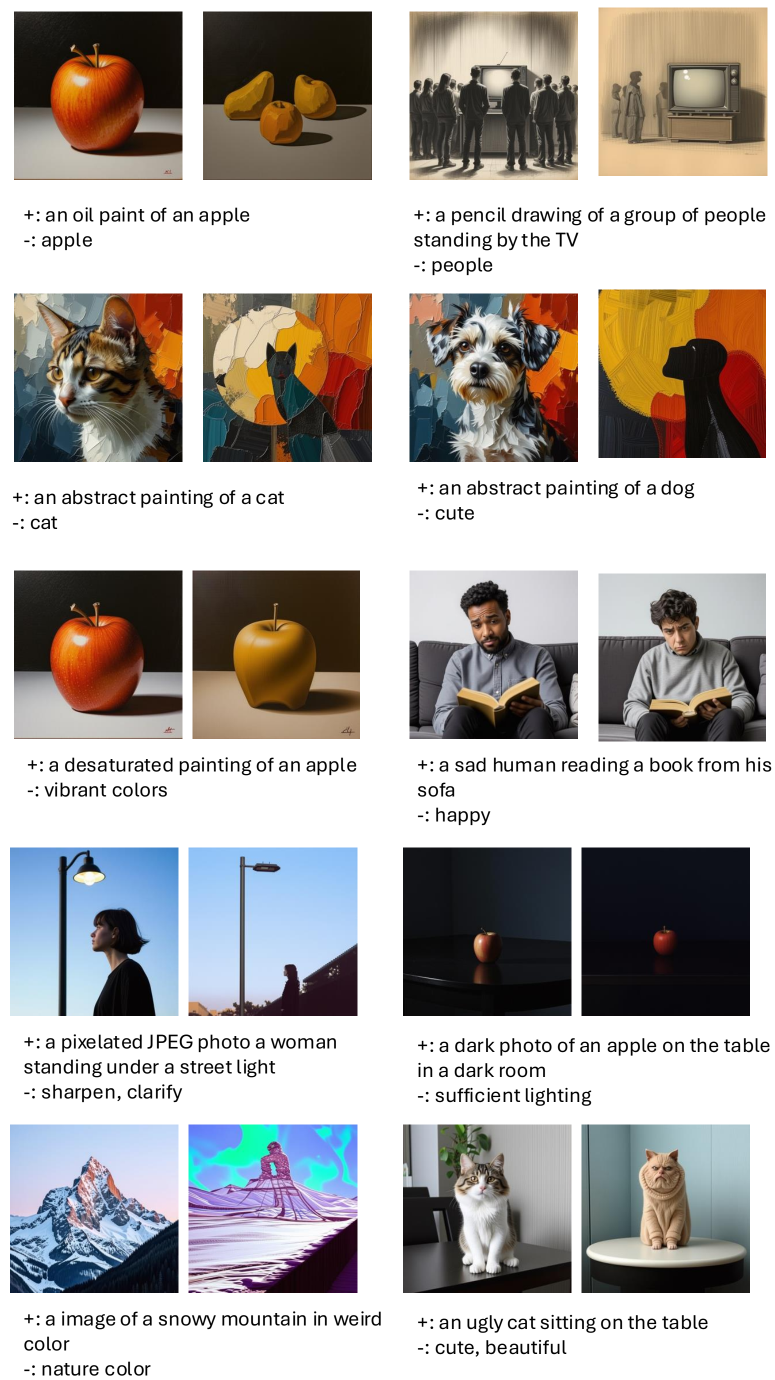

Figure 4: A test for anti-aesthetics. The left image is generated with $\alpha = 0$, and the right image is generated with $\alpha > 0$. These tests aim to move away from universally pleasing styles and demonstrate the ability to capture more diverse aesthetic preferences.

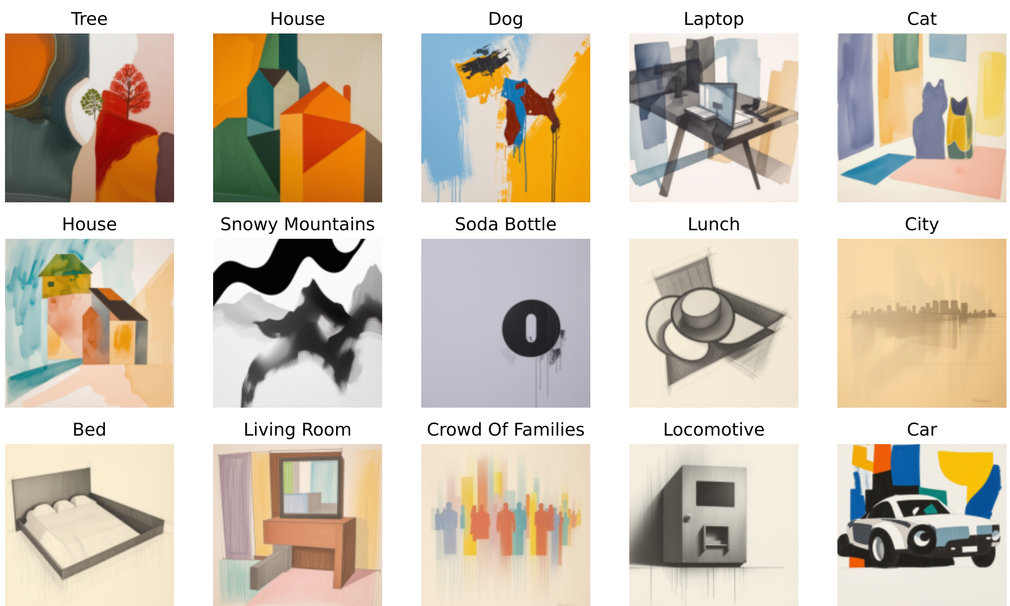

Figure 5: More Abstract Arts Examples