Appendix for *TeEFusion: Blending Text Embeddings to Distill Classifier-Free Guidance*

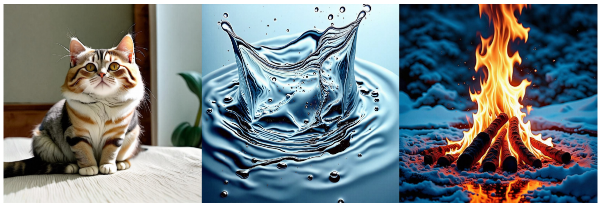

Figure 1. Generation examples of failure cases. Prompt: *1) not a cat. 2) liquid glass. 3) cold fire.*

A. More Experimental Results and Analyses

A.1. Quantitative Analysis of Additive Text Embeddings

To validate the effectiveness of additive text embeddings, we conducted quantitative experiments across different text-to-image models. The cosine similarity between original and fused embeddings (Cos Sim.$_{\text{txt}}$) and their corresponding generated images (Cos Sim.$_{\text{img}}$) are summarized in the table below:

Metric	SD3	In-house T2I	FLUX.1-dev
Cos Sim.$_{\text{txt}}$	0.8073	0.8192	0.8286
Cos Sim.$_{\text{img}}$	0.8732	0.9137	0.9318

These results confirm that additive embedding operations preserve over 80% cosine similarity in text space and over 90% in image space, demonstrating their ability to merge diverse semantic patterns effectively.

A.2. Operational Boundaries and Failure Cases

Our fusion mechanism $\mathcal{G}(\psi(w))\,\mathcal{F}(c - \varnothing)$ operates within the encoder's linear regime through bounded sine-cosine positional encodings ($\|\mathcal{G}(\psi(w))\,\mathcal{F}(c - \varnothing)\|_2 \leq \delta$). However, failure cases arise when:

- Semantic vectors exhibit non-orthogonality (e.g., contradictory phrases like "cold fire")
- Contextual interference occurs in composite prompts (e.g., "not a cat")

These limitations are visualized in Figure 1.