

Target Attribute Diffusion Models

Supplementary Material

9. Additional Experiment Details

9.1. General Configurations

All experiments use Stable Diffusion 2.1 as the base model, which has a resolution of 512 (the default). No quantization was used.

9.2. Experiment 1

This is the single prompt, single attribute experiment meant to demonstrate the basic capabilities of TADM. The prompt is “A photo of a cat” and the attribute is color saturation.

Model hyperparameters The maximum embedding norm (from the attribute cross attention) is 0.1. In the mix encoder, the feedforward module consists of a 3 Linear layer interleaved by SiLU activation functions and layer normalization layers.

Training Adam optimizer with learning rate of 10^{-4} and weight decay of 0.01 was used. During backpropagation, gradients were clipped to a norm of 1.0. The model was trained for exactly 1200 gradient steps with a batch size of 16 without any validation.

9.3. Experiment 2

This is the multi-prompt experiment for three different attributes meant to demonstrate a more general use case of TADM. The three attributes are (a) color saturation, (b) sharpness, and (c) human preference. All experiments for each attribute have the same configuration (besides the attribute measure function):

Model hyperparameters The maximum embedding norm (from the attribute cross attention) is 0.2. In the mix encoder, the feedforward module consists of a 3 Linear layer interleaved by SiLU activation functions and layer normalization layers.

Training Adam optimizer with learning rate of 10^{-4} and weight decay of 0.01 was used. During backpropagation, gradients were clipped to a norm of 0.8. The model was trained for exactly 3000 gradient steps with a batch size of 16. At every 300 steps, TADM generated 4 sample images per prompt and the best checkpoint was saved according to the best alignment score.