UNIVERSITY OF MICHIGAN

Towards Understanding the Mechanism of CFG

Xiang Li, Rongrong Wang, Qing Qu

I. Introduction

Problem: Conditional diffusion models generate images by progressively denoising a random noise $\boldsymbol{x}_T \sim \mathcal{N}\left(\mathbf{0}, \sigma(T)^2 \boldsymbol{I}\right)$ to a clean image with the probabilistic ODE:

$$d\boldsymbol{x} = -\sigma(t)\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|\boldsymbol{c}; \sigma(t))dt, \tag{1}$$

where $\sigma(t)$ is a predefined schedule. However, this standard (naive) sampling approach does not lead to high quality samples, as shown in Figure 1 (left). In contrast, state of the art diffusion models generate high quality images with Classifier-Free-Guidance (CFG):

$$d\boldsymbol{x} = -\sigma(t)\nabla_{\boldsymbol{x}}(\log p(\boldsymbol{x}|\boldsymbol{c}; \sigma(t)dt + \gamma g(\boldsymbol{x}, t)dt), \tag{2}$$

$g(\boldsymbol{x}, t) = \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|\boldsymbol{c}; \sigma(t)) - \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|\sigma(t))$ is the difference between conditional and unconditional scores. With CFG, the quality of conditional generation greatly improves, as shown in Figure 1 (right). This work aims to understand the underlying mechanisms of CFG and we choose linear diffusion models as a prototype since CFG has similar effects on them, as shown in the right part of Figure 1.

II. CFG in Linear Models

In linear diffusion models, CFG guidance $g(\boldsymbol{x}; \sigma(t))$ can be decomposed as:

$$\left(\tilde{\Sigma}_{c,t} - \tilde{\Sigma}_{uc,t}\right)(\boldsymbol{x_t} - \boldsymbol{\mu_c}) + \gamma\left(\boldsymbol{I} - \tilde{\Sigma}_{uc,t}\right)(\boldsymbol{u_c} - \boldsymbol{u_{uc}}). \tag{3}$$

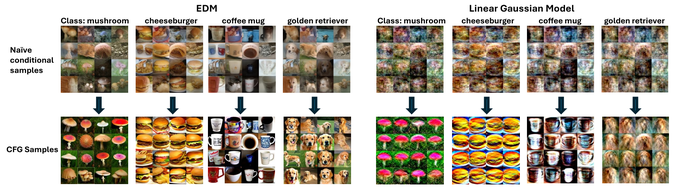

Figure 1: *Left and right parts demonstrate CFG's Effect on nonlinear and linear diffusion models respectively.*

Here:

$$\tilde{\Sigma}_{c,t} - \tilde{\Sigma}_{uc,t} = V_+\Sigma_+V_+ + V_-\Sigma_-V_- \tag{4}$$

, which is the difference between conditional and unconditional data covariance, and it can be decomposed with eigendecomposiiton, separating the positive eigenvector (positive CPC) and negative eigenvector (negative CPC). On the other hand, $\gamma\left(\boldsymbol{I} - \tilde{\Sigma}_{uc,t}\right)(\boldsymbol{u_c} - \boldsymbol{u_{uc}})$ approximately shifts the sample towards the difference between the conditional mean $\boldsymbol{\mu_c}$ and the unconditional mean $\boldsymbol{\mu_{uc}}$.

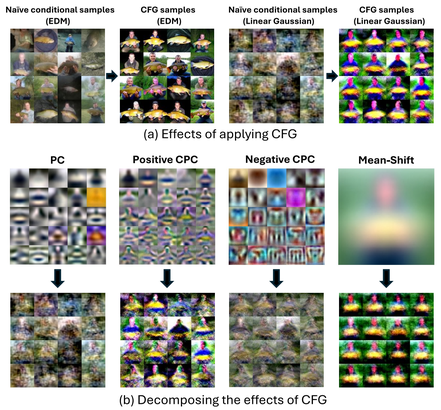

Figure 2: *Distinct effects of different CFG components.*

By decomposing CFG in this way, we can separately examine their distinctive effects, which are shown in Figure 2.
In short, the positive CPC guidance enhances class specific features of the dataset, the negative CPC guidance suppresses conditional-unrelated features in the unconditional dataset and the mean-shift guidance approximately shifts the samples towards the class mean.

III. CFG in Nonlinear Models

For a wide range of noise levels (high noise), diffusion models exhibit linearity and applying linear CFG leads to similar effects as the actual CFG as shown in Figure 3. On the other hand, for low noise levels, diffusion models are highly nonlinear. In this regime, we can construct CPC guidance from the denoiser's Jacobians, which represents the posterior covariances. Such guidance could lead to similar effects as the actual nonlinear CFG (see the paper for more details).

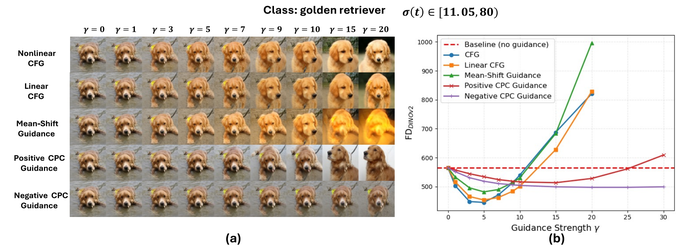

Figure 3: *Distinct effects of different CFG components.*

The main takeaway is that, CFG contrasts between the conditional and unconditional data, highlighting class-specific information while suppressing unrelated ones.

Actionable Interpretability Workshop