Boosting Diffusion Models with Moving Average Sampling in Frequency Domain — CVPR 2024 Supplementary Material*

Yurui Qian†, Qi Cai‡, Yingwei Pan‡, Yehao Li‡, Ting Yao‡, Qibin Sun†, and Tao Mei‡
†University of Science and Technology of China ‡HiDream.ai Inc.
qyr123@mail.ustc.edu.cn, {cqcaiqi, pandy, liyehao, tiyao}@hidream.ai,
qibinsun@ustc.edu.cn, tmei@hidream.ai

This supplementary material contains: 1) the Algorithm underlying our Moving Average Sampling in Frequency domain (MASF); 2) more experiments on ImageNet 64×64; 3) evolution of $\boldsymbol{x}_0^t$ with respect to t; 4) visualization results.

1. Algorithm

We illustrate the algorithm of our MASF in Algorithm S.1.

Algorithm S. 1 MASF

Require: initial value $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \bar{\boldsymbol{x}}_f^{T+1} = \mathbf{0}, \boldsymbol{w}_f^T = \mathbf{0}, (f \in \{ll, lh, hl, hh\})$, variance schedule α_t $(t \in \{1, 2, \ldots, T\}), \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, denoising network $\boldsymbol{\epsilon}_\theta(\cdot)$, solver `Solver`.
MASF hyperparameters: $\gamma, \beta_f(t), \boldsymbol{w}_f^t$.
Denote Discrete Wavelet Transformation as DWT, Inverse DWT as IDWT and element-wise multiplication as $\circ$.
for $t = T$ **to** 1 **do**
  $\boldsymbol{x}_0^t = (\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t))/\sqrt{\bar{\alpha}_t}$
  $\boldsymbol{x}_{ll}^t, \boldsymbol{x}_{lh}^t, \boldsymbol{x}_{hl}^t, \boldsymbol{x}_{hh}^t \leftarrow \text{DWT}\ (\boldsymbol{x}_0^t)$
  for f **in** $\{ll, lh, hl, hh\}$ **do**
    $\bar{\boldsymbol{x}}_f^t \leftarrow (\mathbf{1} - \gamma \boldsymbol{w}_f^t) \circ \boldsymbol{x}_f^t + \gamma \boldsymbol{w}_f^t \circ \bar{\boldsymbol{x}}_f^{t+1}$
  end for
  $\tilde{\boldsymbol{x}}_0^t \leftarrow \text{IDWT}(\beta_{ll}(t)\bar{\boldsymbol{x}}_{ll}^t, \beta_{lh}(t)\bar{\boldsymbol{x}}_{lh}^t, \beta_{hl}(t)\bar{\boldsymbol{x}}_{hl}^t, \beta_{hh}(t)\bar{\boldsymbol{x}}_{hh}^t)$
  $\boldsymbol{x}_{t-1} \leftarrow \texttt{Solver}\ (\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0^t, t)$
end for
return: $\boldsymbol{x}_0$

2. More Experiments on ImageNet 64×64

In this section, we further evaluate our MASF by leveraging three pre-trained models on ImageNet 64×64 dataset in Table S.1 using FID with 50K samples. Generally, applying MASF can boost performances across different models (continuous time and discrete time) and various backbone architectures.

*This work was performed at HiDream.ai.

3. Evolution of $\boldsymbol{x}_0^t$ with respect to t

To validate the efficacy of our MASF on stabilizing denoising process, we further plot the evolution of $\boldsymbol{x}_0^t$ with respect to timestep t before (Baseline) and after applying MASF. Figure S.1 demonstrates that applying MASF can reduce oscillation and lead to a more stable denosing trajectory.

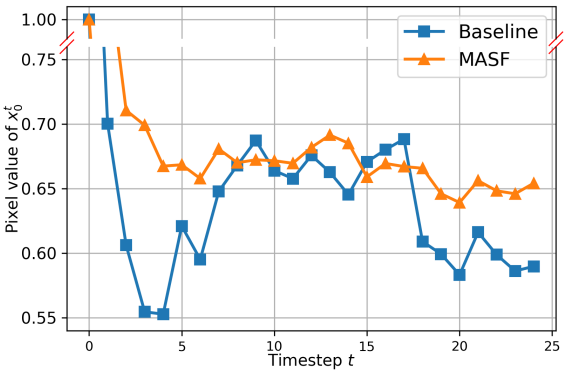

Figure S. 1. We utilize a diffusion model from ADM pre-trained on ImageNet-64 to sample from pure noise and capture the intermediate output as denoised sample $\boldsymbol{x}_t$. We calculate $\boldsymbol{x}_0^t$ with the noise and $\boldsymbol{x}_t$. Subsequently, we plot the pixel value of $\boldsymbol{x}_0^t$ with respect to generative timesteps t from Baseline and our MASF, respectively.

4. Visualization Results

For a more comprehensive qualitative validation of our MASF, we present additional generated images from the ImageNet 256×256 dataset in Figure S.2.

Table S. 1. FID comparisons of 50K samples with different settings on ImageNet 64×64.

Model	Sampler	NFE			
		10	15	20	25
EDM	Heun	33.39	6.12	4.51	3.69
EDM	+MASF	**27.75**	**5.62**	**4.12**	**3.47**
U-ViT	DPM-Solver++(3M)	43.71	5.31	4.41	4.32
U-ViT	+MASF	**29.81**	**4.93**	**4.20**	**3.99**
ADM	DPM-Solver++(2M)	7.34	4.48	3.75	3.49
ADM	+MASF	**6.90**	**4.38**	**3.58**	**3.28**

Figure S. 2. The generated images on ImageNet 256×256 by using DDIM and DDIM plus MASF with 25 NFE (number of function evaluations).