EASING TRAINING PROCESS OF RECTIFIED FLOW MODELS VIA LENGTHENING INTER-PATH DISTANCE

Shifeng Xu[1] Yanzhu Liu[2] Adams Wai-Kin Kong[1]

[1]Nanyang Technological University, Singapore

[2]Institute for Infocomm Research & Centre for Frontier AI Research, A*STAR, Singapore

Outlines

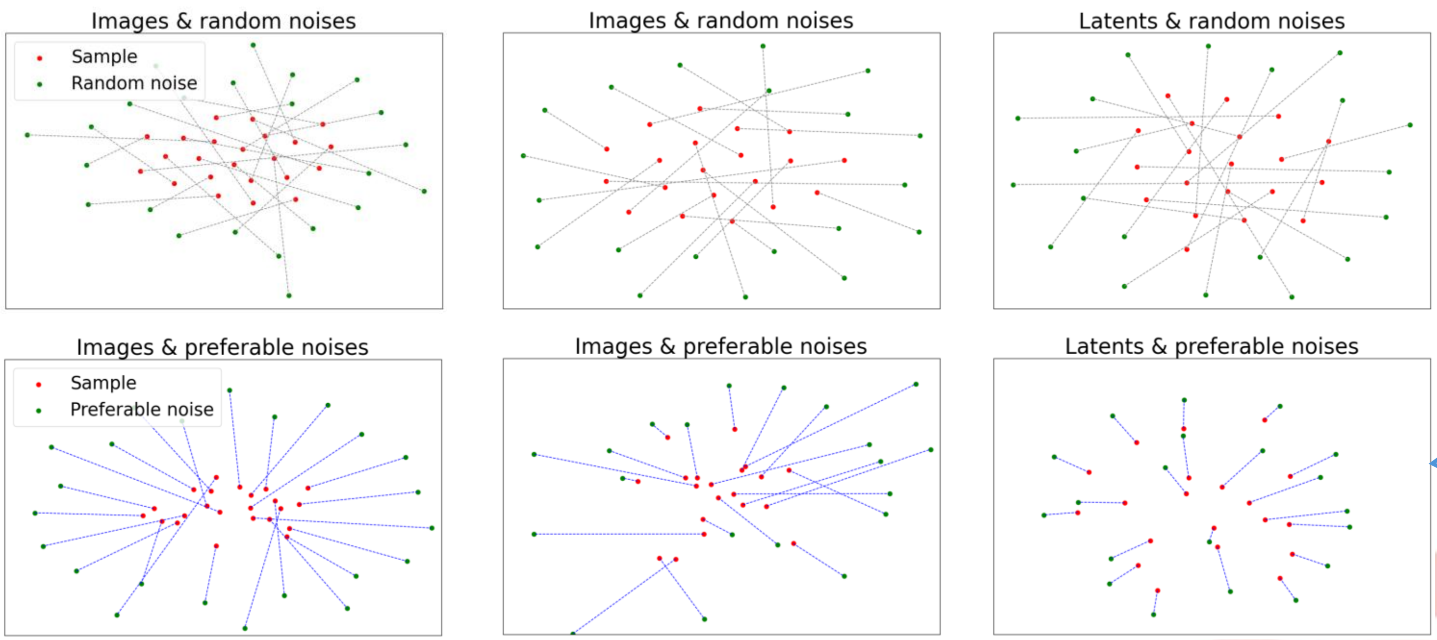
- Problem settings
- Our proposed method: DANSM
- Experiments

Problem settings

- Diffusion-based generative models
  - Diffusion models (DM)
  - Rectified flow models (RFM)
- Consistent model reproducibility
  - From the same noise, generate similar samples

Problem settings

- When converting a sample back to noise,
  - the resultant noise (referred to as **preferable noise**) differs from random noise.
  - The t-SNE visualization of random & preferable noises:

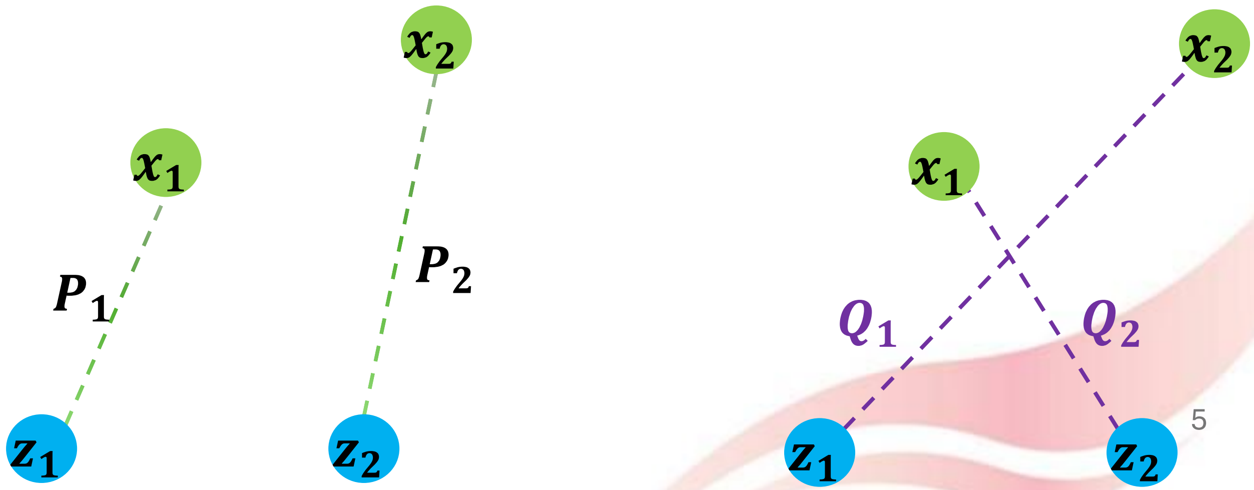


CIFAR-10 images | LSUN Bedroom images | Latents of LSUN Bedroom

The preferable paths have less intersections

Problem settings

- Preferable paths have:
  - **less** intersections in two-dimension visualization,
  - **larger** inter-path distance in high-dimension space.
- in RFM, the paths are straight lines
  - In training process, between 2 samples and 2 noises, which pairs (or paths) are better?



Problem settings

- Problem settings
  - In $\mathbb{R}^{\boldsymbol{d}}$ space, n samples and n noises compose n paths. Let $\boldsymbol{p_{i,j}}$ be the path from the i-th noise to the j-th sample. We define n as the ***match-size***. The objective is to maximize the inter-path distances among these n paths:

$$\max_{\sigma} \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{k=i+1}^{n} dist\left(p_{i,\sigma(i)},\, p_{k,\sigma(k)}\right) ,$$

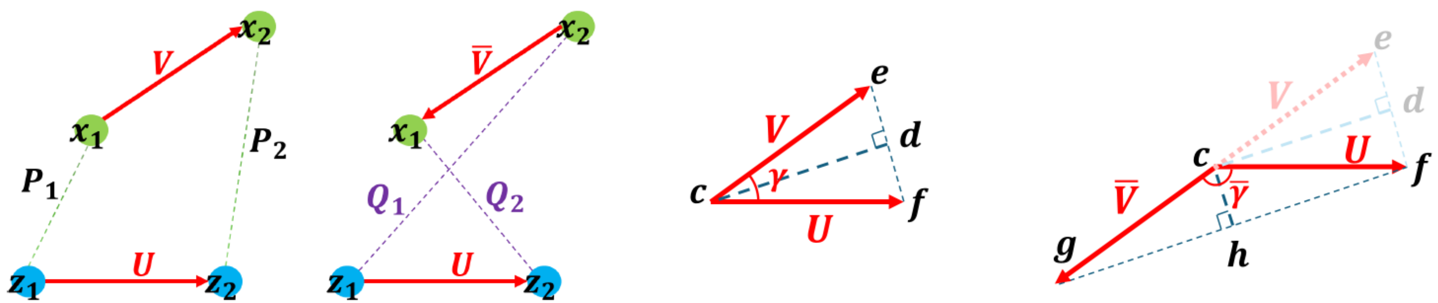
Too high complexity

    where σ is a permutation of $\{1,2,\cdots,n\}$, and $dist(\cdot,\cdot)$ is the distance of two paths.
  - This method is referred to as Distance-Aware Noise-Sample Matching (**DANSM**).

NANYANG TECHNOLOGICAL UNIVERSITY

Our proposed method: DANSM

- We prove that inter-path distance and path length have **negative correlation**.



(a) Case 1 (b) Case 2 (c) Distance in case 1 (d) Distance in case 2

$$\cos\gamma = \frac{\|U\|^2 + \|V\|^2 - \|U - V\|^2}{2\|U\| \cdot \|V\|}$$
$$= \frac{\|Q_1\|^2 + \|Q_2\|^2 - \left(\|P_1\|^2 + \|P_2\|^2\right)}{2\|U\| \cdot \|V\|}$$
$$= -\cos\bar{\gamma}$$

NANYANG TECHNOLOGICAL UNIVERSITY

Shorter path lengths → positive $\cos\gamma$ → acute angle γ → longer inter-path distance

Our proposed method: DANSM

- Surrogate method
  - Instead of lengthening inter-path distance, we **shorten** the path length:

$$\min_{\sigma} \frac{1}{n} \sum_{i=1}^{n} \left\| p_{i,\sigma(i)} \right\|$$

Lower complexity than previous:

$$\max_{\sigma} \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{k=i+1}^{n} dist\left(p_{i,\sigma(i)}, p_{k,\sigma(k)}\right)$$

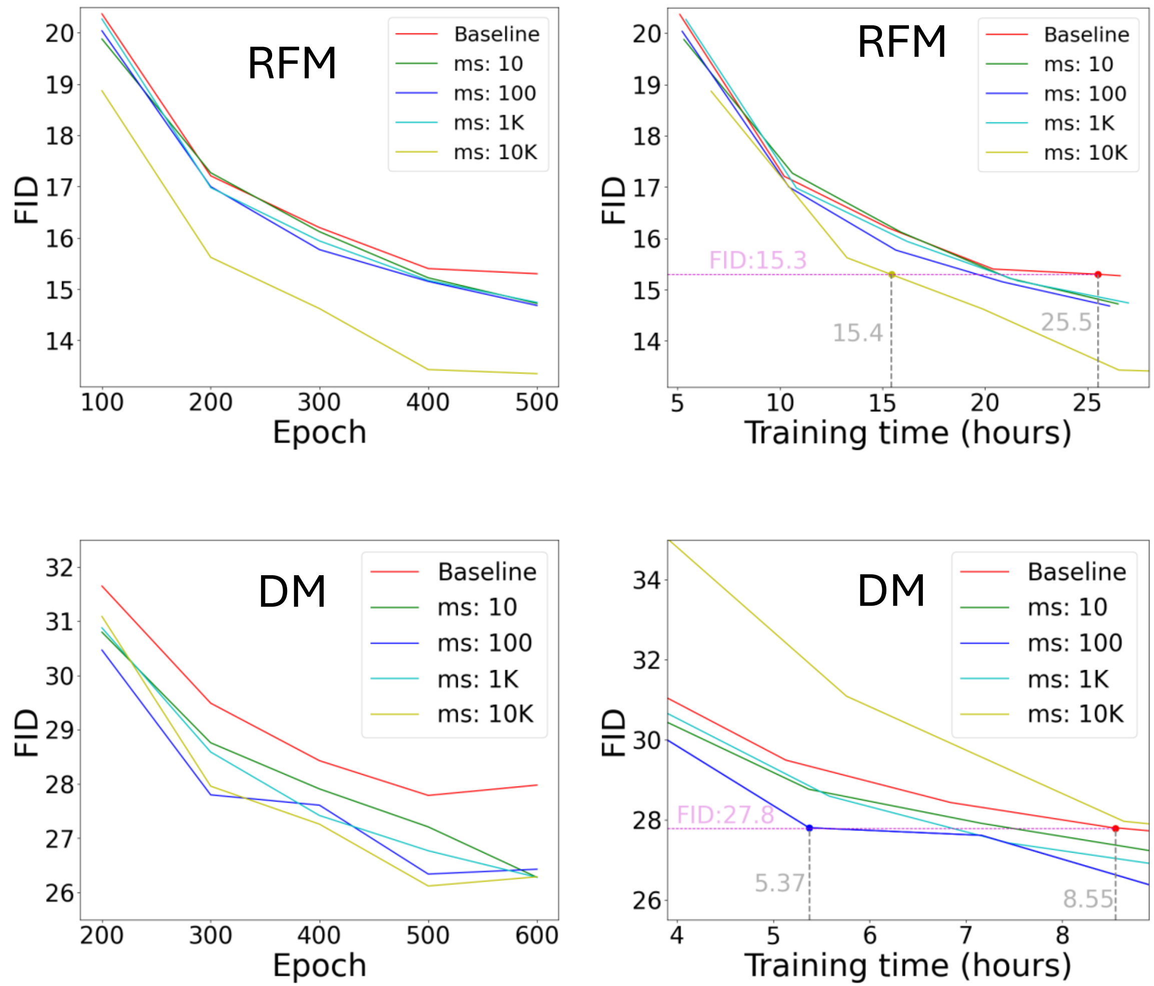
- Implementation
  - Hungarian algorithm has a time complexity of $O(n^3)$.
  - We adopt a **greedy algorithm** with complexity of $O(n^2)$, achieving similar performance.

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments

Training time includes DANSM overhead

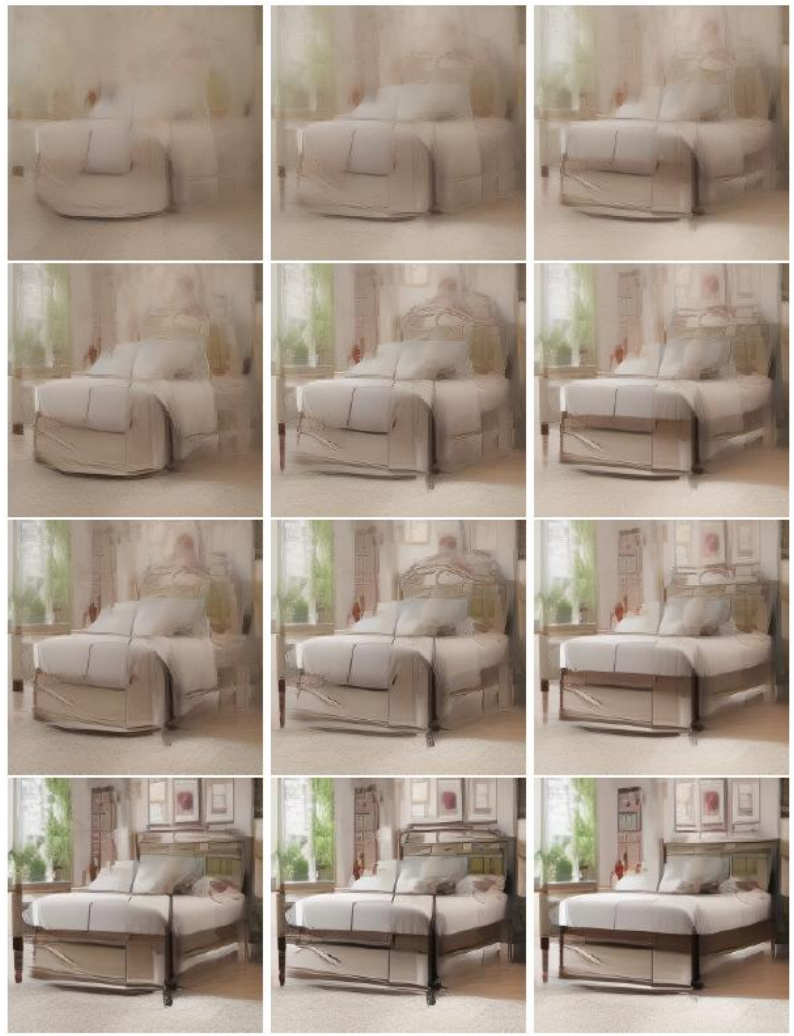
- FID comparison
  - *ms*: match-size
  - match-size n: optimize the noise-sample pairs between n noises and n samples.



NANYANG TECHNOLOGICAL UNIVERSITY

Experiments

- Visualization
  - Samples by
    - 3 steps
    - 4 steps
    - 5 steps
    - 10 steps



Baseline ms:100 ms:10K

Baseline ms:100 ms:1K

NANYANG TECHNOLOGICAL UNIVERSITY

Thanks

NANYANG TECHNOLOGICAL UNIVERSITY