Flow Matching For Generative Modelling

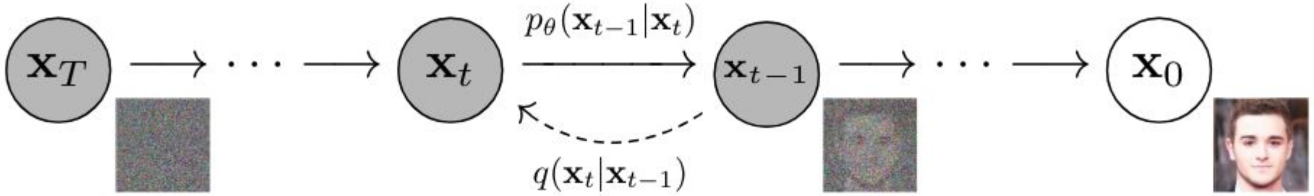
Introduction

- Flow Matching: A new approach for training Continuous Normalizing Flows (CNFs) for generative modeling.

- Key Contributions:
  - The authors depart from the traditional Diffusion and Score matching approaches to generative modeling for image synthesis, instead, the propose a Simulation-free training of CNFs using Flow Matching (FM)
  - The show their work is compatible with various probability paths, including diffusion and Optimal Transport (OT) between distributions.
  - By using OT paths, they show we can achieve faster training, sampling, and better generalization.

Background | Image Synthesis

- Diffusion models: A class of generative models that learn to synthesize images by reversing a gradual noising process
- Formulated as stochastic processes typically based on Gaussian diffusion processes.
- Training objective is denoising score matching: estimates the gradient of the log-density of the noisy data distribution



Forward and reverse diffusion process [1]

- Allows for high-quality image synthesis results and scales to large datasets.
- Confined to simple, hand-designed diffusion processes and slow sampling due to the need for many denoising steps

1. Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Background | Flow Matching

- Flow Matching (FM) offers an efficient, simulation-free approach to train Continuous Normalizing Flows (CNFs) by directly regresses the CNF vector field without explicit maximum likelihood training.
- Key idea: Construct the target vector field by marginalizing simple conditional vector fields
- The CNF vector field is parameterized as a neural network, and gets trained by regressing over the conditional vector fields.
- Sampling is taken care of by solving the ODE associated with the learned neural vector field.

- Key advantages of Flow Matching:
  - Efficient and stable training of CNFs.
  - Avoids the need for explicit maximum likelihood training and expensive simulations.
  - Enables the use of general probability paths, beyond simple diffusion processes.

Problem Definition

- Continuous Normalizing Flows (CNFs) are a continuous-time extension of normalizing flows.
- There are two key components,
  - probability density path: a time dependent probability function, which given a time t and location in space x gives the probability density of that sample

$$p : [0, 1] \times \mathbb{R}^d \to \mathbb{R}_{>0}$$

  - time dependent vector field:

$$v : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$$

- We can thus create a “flow”, defined by an ODE, which captures the rate of change of a sample in the vector field:

$$\frac{d}{dt}\phi(t) = v_t(\phi_t(x)) \longleftarrow \text{Defines the change of flow.}$$

$$\phi_0(x) = x \longleftarrow \text{At t = 0, flow is the point itself}$$

Framework | Flow Matching

- The flow matching objective is to learn a CNF vector field that matches a target vector field corresponding to a desired probability path.
- Given a target probability density path pt(x) and a corresponding vector field ut(x), which generates pt(x), the flow matching objective is:

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t,p_t(x)} \left\| v_t(x) - u_t(x) \right\|^2$$

- While this objective is simple enough, in practice it in intractable since we have no prior knowledge of pt and ut
- The solution is by defining vector fields and paths on a per sample bases (Conditional Flow Matching) with appropriate aggregations leading to pt and pt.

Framework | Flow Matching

- A simple way to construct a target probability path is via a mixture of simpler conditional probability paths.
- Given a sample x1 we can denote a conditional probability path such that:
    - At T=0, P0(x|x1) = p(x), i.e. the source distribution.
    - At T=1, P1(x|x1) is concentrated around x1, (Gaussian around x1 with sufficiently small standard deviation).
- The authors marginalize the conditional probability paths over q(x1) give rise to the marginal probability paths and consequently the can defined the marginal vector fields as well.

$$p_t(x) = \int p_t(x|x_1)q(x_1)dx_1, \quad u_t(x) = \int u_t(x|x_1)\frac{p_t(x|x_1)q(x_1)}{p_t(x)}dx_1,$$

Takeaways: We can break down unknown and intractable marginal VF into simpler conditional VFs, which are much simpler to define as these only depend on a single data sample.

Framework | Conditional Flow Matching

- Unfortunately, due to the intractable integrals in the probability paths and vector fields, it is still intractable to compute ut and pt.
- This is the key contribution from a mathematical perspective, the authors propose Conditional Flow Matching, Unlike FMs the CFM objective allows to easily sample unbiased estimates as long as we can efficiently sample from pt(x|x1)
- The CFM objective is thus defined as:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,q(x_1),p_t(x|x_1)} \left\| v_t(x) - u_t(x|x_1) \right\|^2,$$

- Following this, all we need to do is efficiently define pt and ut on a per sample basis!

Framework | Conditional Probability Paths and Vector Fields

This is where the authors make choices on to fit their problem formulations.

- Firstly, the opt to define pt and ut by a family of Gaussian conditional probability paths.

$$p(x|x_1) = \mathcal{N}(x|\mu_t(x_1), \sigma_t(x_1)^2 I)$$

- Where μ and σ are time dependant mean and standard deviations of the Gaussian.
  - At T=0, the authors set μ0(x1) = 9 and σ(x) = 1, i.e. a standard Gaussian.
  - At T=1, the authors set μ1(x1) = x1 and σ(x) = σmin, such that the distribution is concentrated on x1.
- Using this, the authors define and flow function ψ from which we get the vector field as well.

$$u_t(x|x_1) = \frac{\sigma_t'(x_1)}{\sigma_t'(x_1)}(x - \mu_t(x_1)) + \mu_t'(x_1)$$

Framework | Diffusion Conditional Vector Fields

- CFMs can be viewed as a generalisation of the diffusion process with specific selection of pt and ut.
- A special case to note is that of the Variance Preserving Diffusion process where pt and ut are given by:

$$p_t(x|x_1) = \mathcal{N}(x \mid \alpha_{1-t}x_1, \left(1 - \alpha_{1-t}^2\right) I)$$

$$u_t(x|x_1) = \frac{\alpha'_{1-t}}{1-\alpha_{1-t}^2}\left(\alpha_{1-t}x - x_1\right) = -\frac{T'(1-t)}{2}\left[\frac{e^{-T(1-t)}x - e^{-\frac{1}{2}T(1-t)}x_1}{1-e^{-T(1-t)}}\right]$$

- In generating the vector field with u1, at t = T, the field is undefined, and thus would never reach a true noise distribution, we simply assume it approximates one since we can't do infinite steps.
- In contrast, CFMs only need finite steps!

Framework | Optimal Transport Vector Fields

- The authors offer a much simpler choice of pt and ut, a simple linear time interpolation between the two distributions.

$$\mu_t(x) = tx_1; \text{ and } ;\sigma_t(x) = 1 - (1 - \sigma_{min})t$$

- Which yields a vector field which is constant in direction over time, unlike diffusion.

$$u_t(x|x_1) = \frac{x_1 - (1 - \sigma_{min})x}{1 - (1 - \sigma_{min})t}$$

- This linear definition means particles under the OT displacement map always move in straight line trajectories and with constant speed

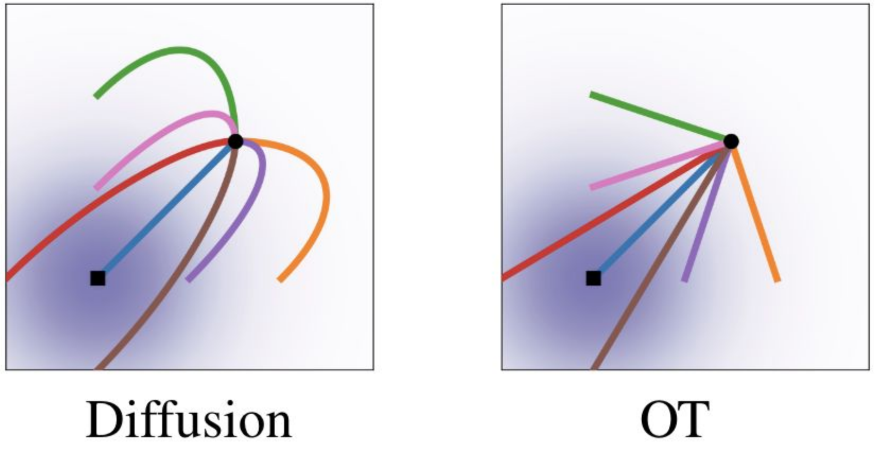
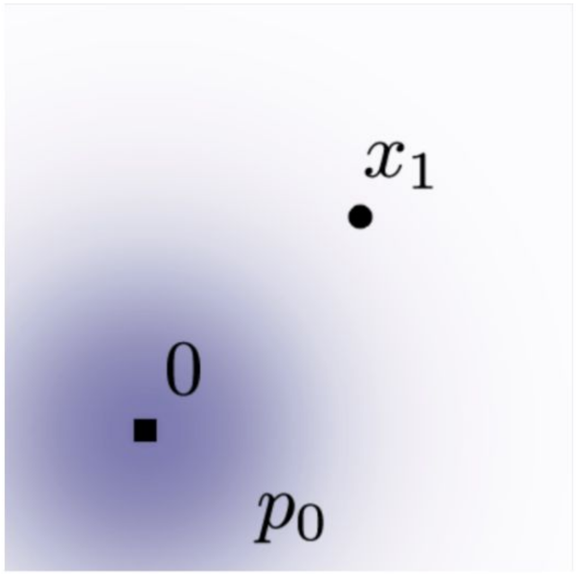
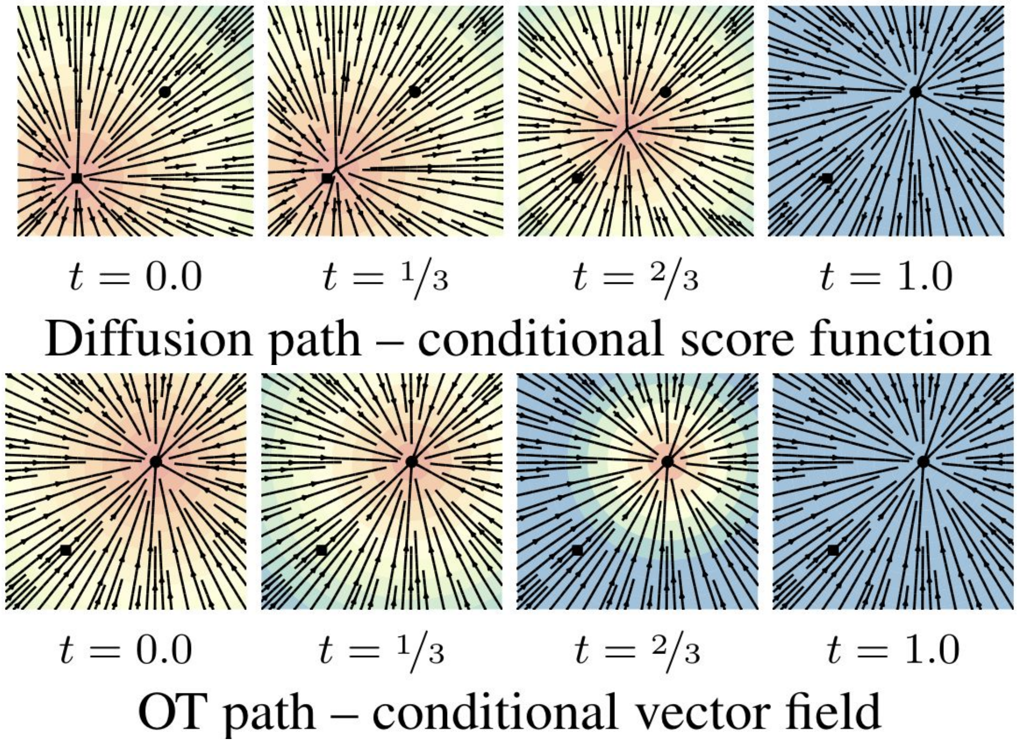


Figure 3: Diffusion and OT trajectories.

Framework | Optimal Transport vs Diffusion Paths

- The diffusion vector follow the diffusion process, which change over time in direction and scale.
- In contrast the Optimal transport paths are constant.





Experiments

- Comparing Flow Matching (FM) with probability paths based on Diffusion & Optimal Transport (OT) to models with Diffusion based losses and GANs

Model	CIFAR-10 NLL↓	CIFAR-10 FID↓	CIFAR-10 NFE↓	ImageNet 32×32 NLL↓	ImageNet 32×32 FID↓	ImageNet 32×32 NFE↓	ImageNet 64×64 NLL↓	ImageNet 64×64 FID↓	ImageNet 64×64 NFE↓
Ablations									
DDPM	3.12	7.48	274	3.54	6.99	262	3.32	17.36	264
Score Matching	3.16	19.94	242	3.56	5.68	178	3.40	19.74	441
ScoreFlow	3.09	20.78	428	3.55	14.14	195	3.36	24.95	601
Ours									
FM ^{w}/ Diffusion	3.10	8.06	183	3.54	6.37	193	3.33	16.88	187
FM ^{w}/ OT	**2.99**	**6.35**	**142**	**3.53**	**5.02**	**122**	**3.31**	**14.45**	**138**

Model	ImageNet 128×128 NLL↓	ImageNet 128×128 FID↓
MGAN (Hoang et al., 2018)	–	58.9
PacGAN2 (Lin et al., 2018)	–	57.5
Logo-GAN-AE (Sage et al., 2018)	–	50.9
Self-cond. GAN (Lučić et al., 2019)	–	41.7
Uncond. BigGAN (Lučić et al., 2019)	–	25.3
PGMGAN (Armandpour et al., 2021)	–	21.7
FM ^{w}/ OT	**2.90**	**20.9**

Table 1: Likelihood (BPD), quality of generated samples (FID), and evaluation time (NFE) for the same model trained with different methods.

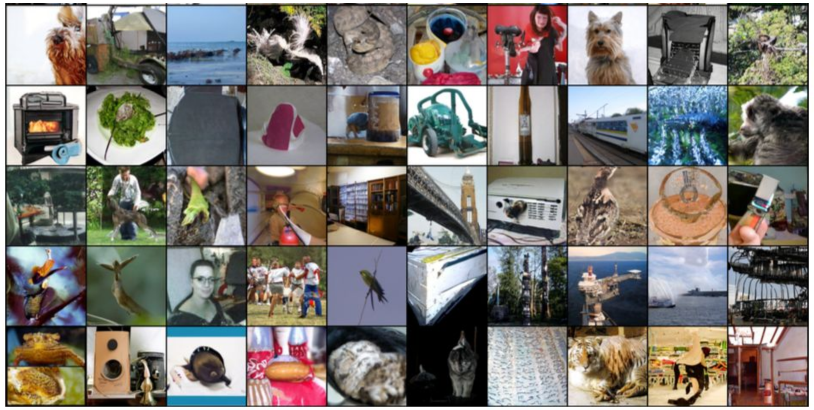


Figure 13: Non-curated unconditional ImageNet-128 generated images of a CNF trained with FM-OT.

Experiments

- Flow Matching results in faster training
- For conditional image generation - upsampling images from 64x64 to 256x256 - FM with OT improves FID score

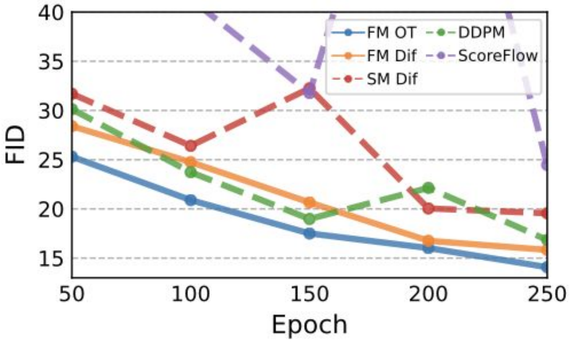


Figure 5: Image quality during training, ImageNet 64×64.

Conditional Image Generation - Upsampling

Model	FID↓	IS↑	PSNR↑	SSIM↑
Reference	1.9	240.8	–	–
Regression	15.2	121.1	**27.9**	**0.801**
SR3 (Saharia et al., 2022)	5.2	180.1	26.4	0.762
FM ^{w}/ OT	**3.4**	**200.8**	24.7	0.747

Table 2: Image super-resolution on the ImageNet validation set.

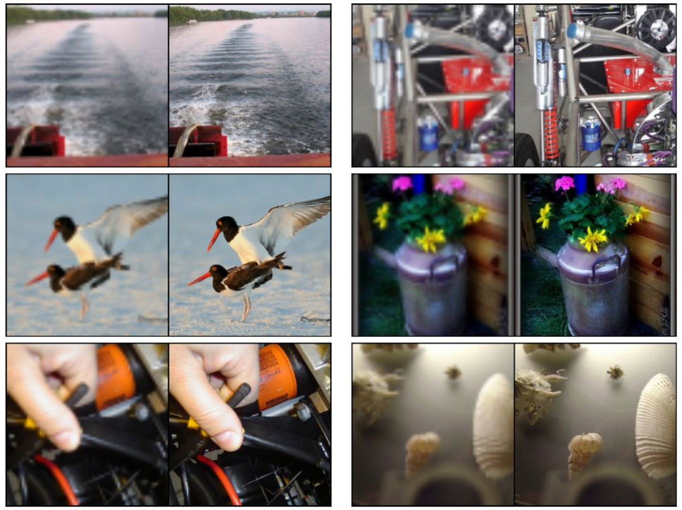
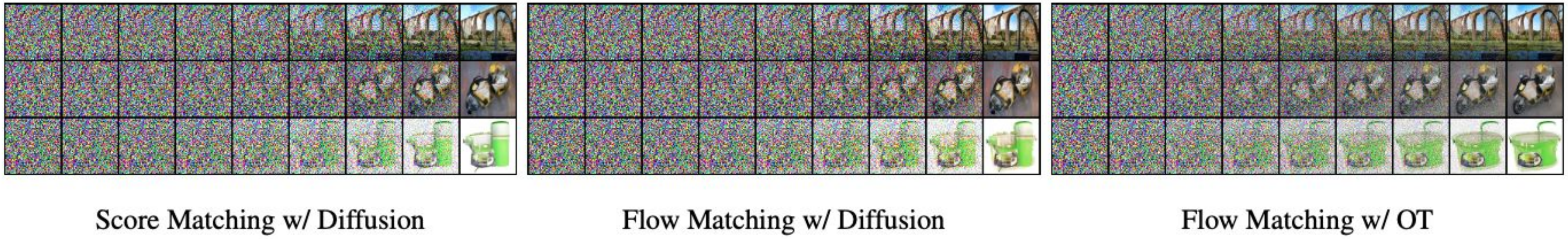


Figure 14: Conditional generation 64×64→256×256. Flow Matching OT upsampled images from validation set.

Experiments

- Flow Matching with OT path start generating images sooner than diffusion based models



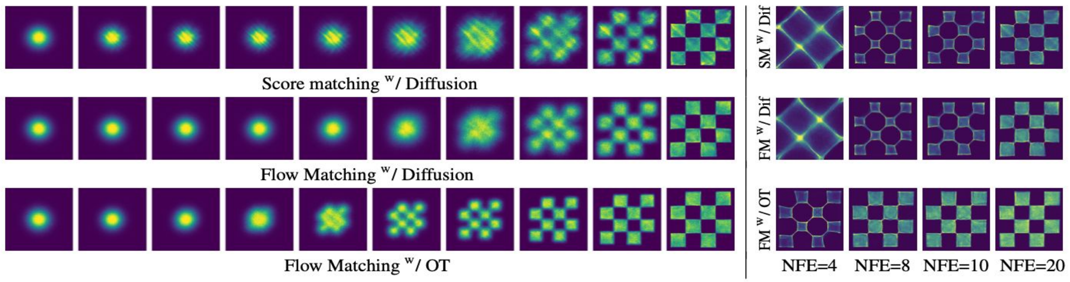


Figure 4: (*left*) Trajectories of CNFs trained with different objectives on 2D checkerboard data. The OT path introduces the checkerboard pattern much earlier, while FM results in more stable training. (*right*) FM with OT results in more efficient sampling, solved using the midpoint scheme.

Experiments

- When using fixed step solvers & comparing low (<= 100) NFE samples with samples from 1,000 NFE solutions through per pixel MSE calculation → FM with OT produces lowest error
- FM with OT achieves decent FID even at low NFE values → better trade-off between sample quality & cost compared to diffusion based models

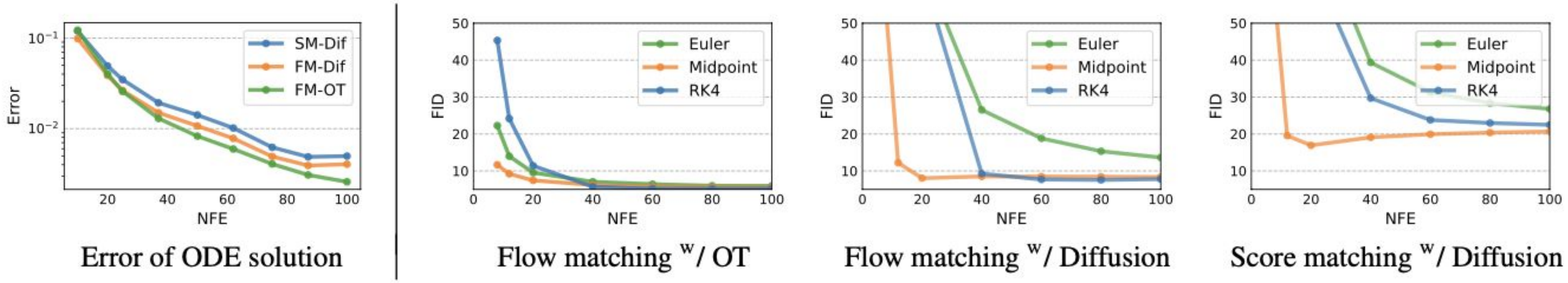


Figure 7: Flow Matching, especially when using OT paths, allows us to use fewer evaluations for sampling while retaining similar numerical error (left) and sample quality (right). Results are shown for models trained on ImageNet 32×32, and numerical errors are for the midpoint scheme.

Conclusion

- Conditional Flow Matching For CNFs enables scaling CNFs to very high dimensions
- Flow Matching offers simulation-free approach
- By directly specifying probability paths it allows faster sampling and improved image generation