Multisample Flow Matching: Straightening Flows with Minibatch Couplings

Aram-Alexandre Pooladian*, Heli Ben-Hamu*, Carles Domingo-Enrich*, Brandon Amos, Yaron Lipman, Ricky T. Q. Chen
(*equal contribution)
Meta AI
NEW YORK UNIVERSITY
WEIZMANN INSTITUTE OF SCIENCE

ICML International Conference On Machine Learning

Introduction

Simulation-free continuous-time generative models have achieved state of the art performance across various data modalities. Although showing high sample quality, these models require a costly sampling procedure consisting of many function evaluations. Recently, [1] proposed the Flow Matching framework for training continuous normalizing flows in a simulation-free manner and introduced the conditional optimal transport (OT) probability paths, which improved sampling speed compared to diffusion paths.

Goals

- **Extend** the Flow Matching framework for **general joint distributions**, $q(x_0, x_1)$.
- Find joint distributions $q(x_0, x_1)$ that simplify ODE trajectories — **faster sampling**.

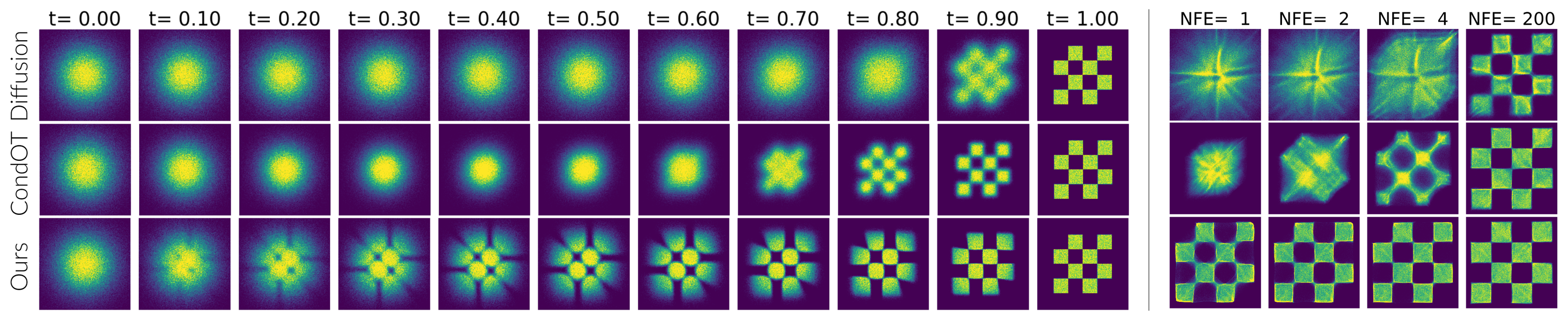

Figure 1. Probability paths (left) and sampling with limited NFE (right). Ours ran with BatchOT couplings.

Preliminaries

Continuous Normalizing Flow (CNF). A time-dependent diffeomorphic map, $\psi : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$, parameterized by the vector field $u_t(x)$ and defined via the ordinary differential equation:

$$\frac{d}{dt}\psi_t(x_0) = u_t(\psi_t(x_0)), \quad \psi_0(x_0) = x_0. \tag{1}$$

A CNF induces a continuous transformation between probability densities defined by the push-forward $p_t = \psi_{t*}q_0$ where in generative modeling we seek to approximate some target distribution q_1, that is, we want: $p_1 \approx q_1$.

Flow Matching. A simulation-free method to train CNFs. The core idea is to implicitly construct a probability path, $p_t(x)$, between the source distribution q_0 and the target distribution q_1 using a conditional probability path that is easy to sample from:

$$p_t(x) = \int p_t(x|x_1)q_1(x_1)dx_1, \quad p_0(x|x_1) = q_0(x), \quad p_1(x|x_1) = \delta(x - x_1),$$

The marginal vector field, $u_t(x)$, generating $p_t(x)$ is the minimizer of the *Conditional Flow Matching* loss:

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{t,q_0(x_0),q_1(x_1)} \left\| v_t(x_t;\theta) - u_t(x_t|x_1) \right\|^2,$$

where $u_t(x|x_1)$ is the conditional vector field generating $p_t(x|x_1)$, and $x_t = \psi_t(x_0|x_1)$ is the corresponding conditional transport map. In particular, [1] proposes an OT inspired conditional path called *CondOT* for which $u_t(x|x_1) = x_1 - x_0$.

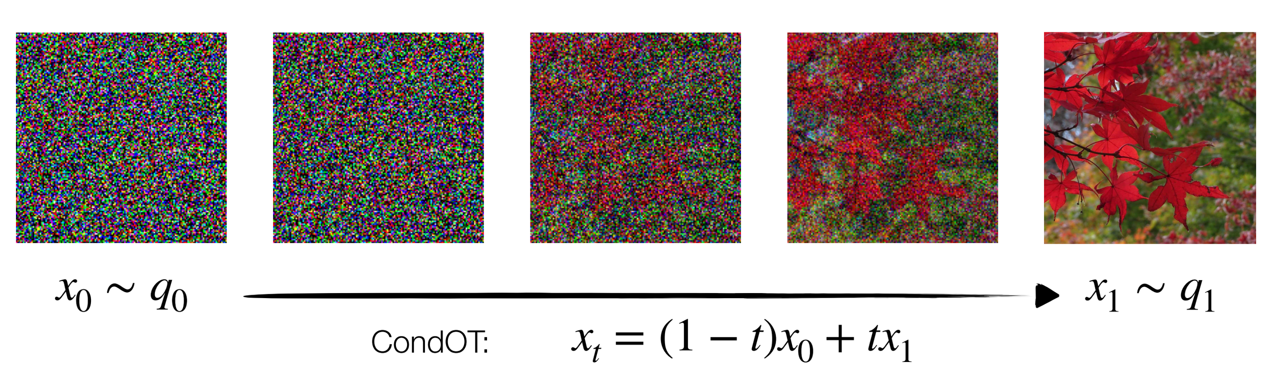

Figure 2. Conditional probability path samples x_t for the CondOT path.

tl;dr

1. Multisample FM extends the family of probability paths used for simulation-free training of CNFs.
2. Using minibatch solutions of optimal transport, we achieve:
   - Faster convergence during training.
   - Faster sampling.
3. First work capable of utilizing minibatch optimal transport couplings to produce a marginal preserving map between distributions.

Multisample Flow Matching

Joint Flow Matching. We show that for a joint distribution $q(x_0, x_1)$ satisfying:

$$\int q(x_0, x_1)dx_1 = q_0(x_0), \quad \int q(x_0, x_1)dx_0 = q_1(x_1),$$

minimizing the *Joint Conditional Flow Matching* loss

$$\mathcal{L}_{\text{JCFM}} = \mathbb{E}_{t,q(x_0,x_1)} \left\| v_t(x_t;\theta) - u_t(x_t|x_1) \right\|^2,$$

yields the marginal vector field, $u_t(x)$, that pushes q_0 to q_1, similarly to the CFM objective.

Multisample Flow Matching. The JCFM loss adds another degree of freedom in the design space of probability paths. However, it is non-trivial to construct such joint distributions satisfying the marginals. We propose a constructive approach towards this goal, called *Multisample Flow Matching*.

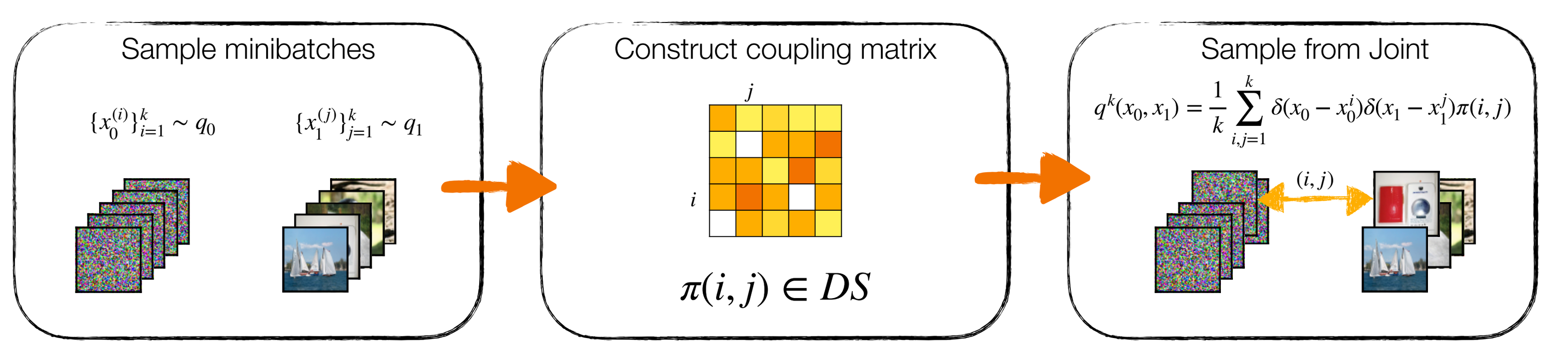

Figure 3. **The Multisample Flow Matching Algorithm.** We randomly sample noise and data samples, then re-arrange the pairing to be either optimal, or stable, within the current minibatch.

Batch Optimal Transport Couplings

Optimal Transport.

$$W_2^2(q_0, q_1) = \underbrace{\min_{q \in \Gamma(q_0,q_1)} \mathbb{E}_{q(x_0,x_1)}[\|x_0 - x_1\|^2]}_{\text{Static Cost}} = \underbrace{\min_{p_t,u_t} \int_0^1 \int_{\mathbb{R}^d} \|u_t(x)\|^2 p_t(x)\mathrm{d}x\mathrm{d}t}_{\text{Dynamic Cost}},$$

Aiming to simplify the ODE trajectories of the marginal vector field, we propose to use minibatch OT couplings, denoted as *BatchOT*, to construct $\pi(i,j)$.

Theorem (informal). Suppose that Multisample Flow Matching is run with BatchOT. Then, as $k \to \infty$,

1. The value of the JCFM objective for the optimal u_t converges to 0.
2. A straightness measure, S, of the optimal u_t converges to zero.
3. The static cost associated to the optimal u_t is monotonically decreasing and converges to the OT cost $W_2^2(p_0, p_1)$.

	CondOT	BatchOT	BatchEOT	Stable	Heuristic
Runtime Complexity	$\mathcal{O}(1)$	$\mathcal{O}(k^3)$	$\tilde{\mathcal{O}}(k^2/\epsilon)$	$\mathcal{O}(k^2\log(k))$	$\mathcal{O}(k^2\log(k))$

Table 1. Runtime complexities of the **alternative** coupling algorithms as a function of the batch size k.

Experiments

We experiment with large-scale image datasets ImageNet32 and ImageNet64. We show two main contributions: (i) faster convergence at training and (ii) better image quality at fixed compute budgets, measured using number of function evaluations (NFEs).

Faster Convergence and Sampling.

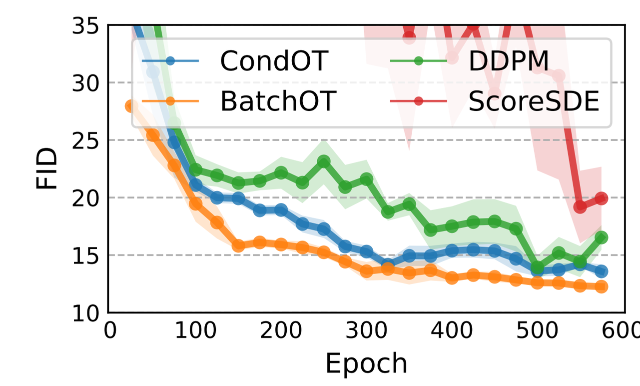

Figure 4. Multisample Flow Matching with BatchOT shows faster convergence on (ImageNet64).

	ImageNet 32×32 NFE @ FID $= 10$	ImageNet 64×64 NFE @ FID $= 20$
Diffusion	≥40	≥40
FM w/ CondOT	20	29
MultisampleFM w/ Heuristic	18	12
MultisampleFM w/ Stable	**14**	**11**
MultisampleFM w/ BatchOT	**14**	12

Table 2. NFE required to achieve a certain FID across our proposed methods. The baseline diffusion-based methods (e.g., ScoreFlow and DDPM) require more than 40 NFE to achieve these FID values. **Surprisingly, our Stable couplings performs on par with BatchOT couplings at lower compute.**

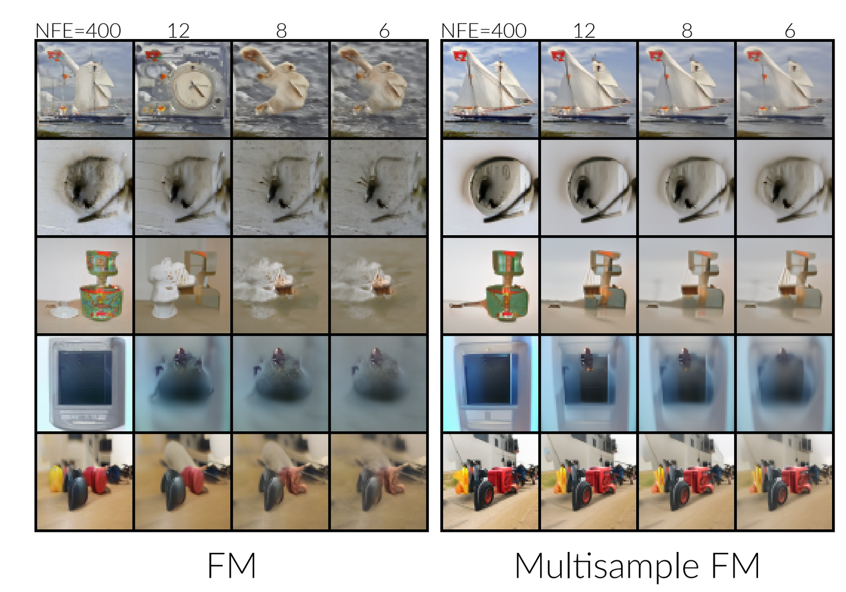

Figure 5. Multisample Flow Matching trained with BatchOT couplings produces more consistent samples across varying NFEs. Note that both flows on each row start from the same noise sample (ImageNet64).

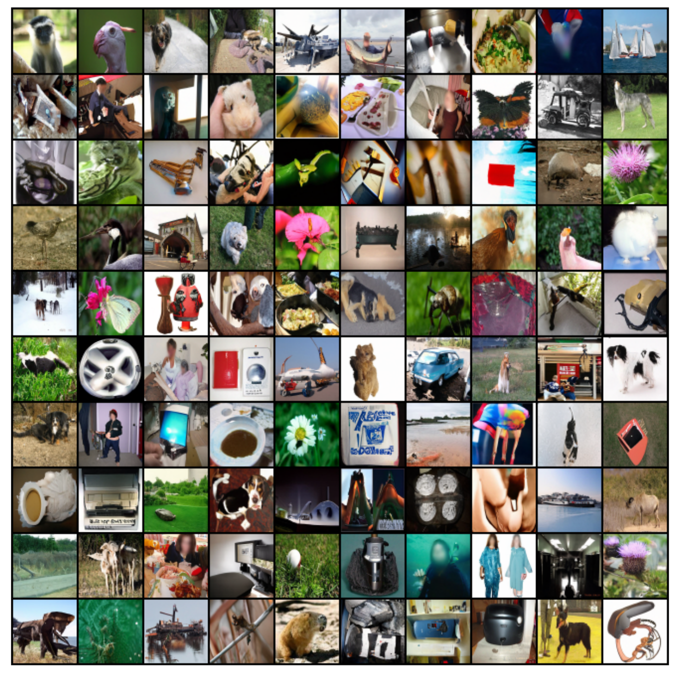

Figure 6. Non-curated samples from a model trained on ImageNet64.

Transport Cost vs. Batch Size

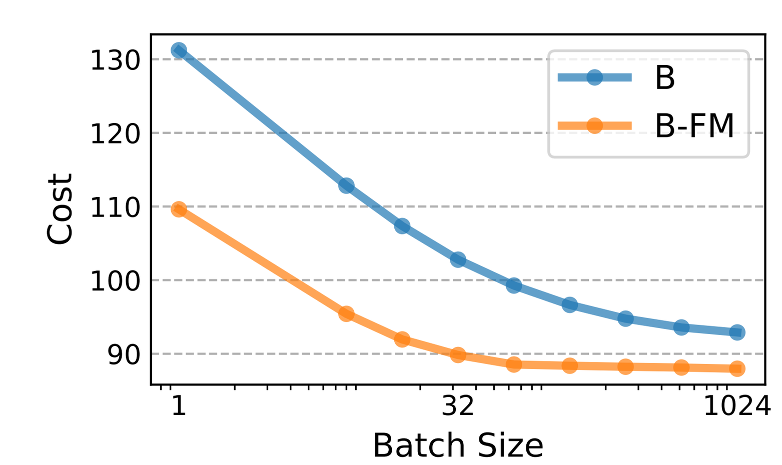

Figure 7. Flow Matching reduces the transport cost further, compared to minibatch OT solutions. Ablation on the static transport cost as a function of the batch size k in $\mathbb{R}^{64}$. Blue: cost of minibatch OT solutions, Orange: cost of trained flow with BatchOT multisample FM.

References

[1] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *International Conference on Learning Representations*, 2023.