On Conditional Sampling with Joint Flow Matching

Amy Xiang Wang [1]

Abstract

A transport map is versatile and useful for many downstream tasks, from training generative modeling to solving Bayesian inference problems. (Marzouk et al., 2016) pioneered the *measure transport* approach for sampling by introducing its connection with transport map $T_\sharp \rho = \mu$, where samples from μ can be easily drawn. When the transport map is a lower-triangular map or *Knothe-Rosenblatt map*, we can also draw conditional samples $\mu_{2|1}(\cdot|x_1)$ from the target distribution μ generalized by (Kovachki et al., 2020). This state-of-the-art sampling approach deviates from traditional methods such as MCMC or variational inference and has received many research interests in recent years. In our work, we introduce a new approach to approximate this transport map to perform conditional sampling tasks using a recent computational advance in generative modeling – flow matching. Specifically, we use (Pooladian et al., 2023a)'s joint flow matching approach with a twisted Euclidean cost to ensure the triangular property of the map. We empirically verify our method through benchmark examples and quantifying the approximated map errors.

1. Introduction

Conditional sampling for generative models is widely applicable in many areas from image editing to super resolution medical imaging. Indeed, generative models such as diffusion models are accredited for many of these real-world applications for their impressive unconditional generation abilities, but altering these models from unconditional to conditional can be challenging. In this paper, we examine from a different perspective and connect two frameworks to introduce a simple and elegant way for conditional sampling. We tackle this task mainly from the measure transport perspective and merge two frameworks: joint flow matching and optimal transport with a twisted cost function. In particular, we build upon the following result from (Kovachki et al., 2020) which proved a generalized version of (Marzouk et al., 2016) lemma one.

$$T^{(2)}(x_1, \cdot)_\sharp \rho_2 = \mu_{2|1}(\cdot|x_1) \tag{1}$$

where $T : \mathbb{R}^{d_1+d_2} \to \mathbb{R}^{d_1+d_2}$ is a transport map between two probability measures ρ and μ, over $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, written $T_\sharp \rho = \mu$, if for $X \sim \rho$, then $T(X) \sim \mu$, and $T^{(2)}(x_1, \cdot) : \mathbb{R}^{d_2} \to \mathbb{R}^{d_2}$ for every fixed $x_1 \in \mathbb{R}^{d_1}$ is the second component of the map.

Estimating this transport map or conditional Brenier map can be daunting, especially in high-dimensional settings. Here, we utilize joint flow matching (Pooladian et al., 2023a), a recent advancement in generative modeling, and the connection between the conditional Brenier map and another type of transport map: Knothe-Rosenblatt map or KR map to estimate the latter one instead. The KR map is a lower triangular lexicographic order map. It is much easier to estimate as it is essentially one-dimensional, easy to construct with explicit formula,and its structure is well-suited for conditional sampling tasks. We can do so by using the optimal transport theory and a twisted cost function to impose the KR map structure (Carlier et al., 2010).

2. Background

We provide only a bare-bones introduction to optimal transport; we recommend (Santambrogio, 2015; Villani, 2009) for more information on this topic.

Optimal transport Let $\rho, \mu \in \mathcal{P}(\mathbb{R}^d)$ be two probability measures in $\mathbb{R}^d$. Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function of the form $c(x, y) := h(x - y)$ for $h : \mathbb{R}^d \to \mathbb{R}$ strictly convex. The Wasserstein distance for the cost c is given by the Kantorovich formulation (Kantorovitch, 1942)

$$W_c(\rho, \mu) := \min_{\pi \in \Gamma(\rho,\mu)} \iint c(x, y) d\pi \,, \tag{2}$$

where $\Gamma(\rho, \mu) \subseteq \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is the set of couplings between ρ and μ: the first marginal of π is ρ, and the second is μ. When ρ and μ have finite second-moment, a minimizer to W_c always exists, called the *optimal coupling*, and is

[*]Equal contribution [1]Courant Institute of Mathematical Sciences, Department of Computer Science, New York University. Correspondence to: Amy Xiang Wang <xw914@nyu.edu>.

Accepted by the Structured Probabilistic Inference & Generative Modeling workshop of ICML 2024, Vienna, Austria. Copyright 2024 by the author(s).

denoted by π_c. Equation 2 is known as the *primal* optimal transport problem associated to the cost function c.

Conditional Brenier map Let $\rho, \mu \in \mathcal{P}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$. Suppose $\rho_1, \mu_1 \in \mathcal{P}(\mathbb{R}^{d_1})$, and ρ_1 has a density. Let $\nabla\phi^{(1)} : \mathbb{R}^{d_1} \to \mathbb{R}^{d_1}$ be the optimal transport map for the quadratic cost $c(x, y) := \frac{1}{2}\|x - y\|^2$ between ρ_1 and μ_1, where ϕ is a convex function called an optimal Brenier potential. Now, for any $x_1 \in \operatorname{supp}(\rho_1)$, consider $\rho_{2|1}(\cdot|x_1)$ and $\mu_{2|1}(\cdot|x_1)$, both in $\mathcal{P}(\mathbb{R}^{d_2})$, and again assume the source measure $\rho_{2|1}(\cdot|x_1)$ has a density. Now, define $\nabla\phi^{(2)}(x_1, \cdot) : \mathbb{R}^{d_2} \to \mathbb{R}^{d_2}$ to be the optimal transport map for the squared-Euclidean cost between these two measures. The conditional Brenier map is defined as

$$T_{\text{CB}}(x_1, x_2) := [\nabla\phi^{(1)}_{\text{CB}}(x_1); \nabla\phi^{(2)}_{\text{CB}}(x_1, x_2)]\,. \tag{3}$$

Knothe-Rosenblatt map Suppose $d_1 + d_2 = d$; we can recursively apply the definition of Equation 3 until we obtain d conditional measures with d conditional maps. Stacking the maps as before, this results in the *Knothe-Rosenblatt* map

$$\begin{aligned} T_{\text{KR}}(x_1, \ldots, x_d) = [\nabla\phi^{(1)}_{\text{KR}}(x_1); \nabla\phi^{(2)}_{\text{KR}}(x_1, x_2); \cdots ; \\ \nabla\phi^{(d)}_{\text{KR}}(x_1, x_2, \ldots, x_d)]\,. \end{aligned} \tag{4}$$

Note that unlike the conditional Brenier maps, each component is univariate. See more in section 3.

2.1. Flow Matching

Flow matching (Lipman et al., 2022; Albergo et al., 2023; Liu et al., 2022) belongs to the family of continuous normalizing flow (CNF) (Chen et al., 2022) – a class of generative model approach where a simple prior distribution such as standard Gaussian flow along a neural network trained vector field v_t to the target data distribution. (Lipman et al., 2022) used individual data samples to construct the probability path. Specifically, it used conditional probability to simplify a previously intractable learning objective for the unknown pair: vector field v_t and probability path p_t generated by u_t through the following:

$$\min_\theta \int_0^1 \iint \|v_\theta(t, x_t) - u_t(x_t|y)\|_2^2 d\rho_t(x_t|y) d\mu(y) dt\,, \tag{5}$$

where $\rho_t(\cdot|y)$ and $u_t(\cdot|y)$ are known by design.

Joint Straight Flow Matching (JSFM) (Pooladian et al., 2023a; Tong et al., 2023) further build on (Lipman et al., 2022)'s flow matching by introducing a more general framework with joint distribution and mini-batches training called Multisample Flow Matching or OT-CFM. The key improvements of this approach are as follows: it decreases the computational cost of flow matching by *inducing* optimality in the trajectories, resulting in fewer evaluation calls to the fitted neural network to generate samples. The optimal transport path applied here constructs a straighter path. The training objective is as follows:

$$\min_\theta \int_0^1 \iint \|v_\theta(t, x_t) - (y - x)\|_2^2 dq(x, y) dt\,, \tag{6}$$

3. Conditional Sampling with JSFM

In this section, we introduce our approach to drawing conditioned samples using Joint Straight Flow Matching by introducing a twisted Euclidean cost, the algorithm, and discussing the advantage of using KR rearrangement for estimating the transport map.

3.1. Twisted Euclidean Cost

The conditional Brenier map and the Knothe-Rosenblatt map are two types of transport maps. The former map is challenging to approximate in high dimensional settings but crucial as it is the optimal transport map to solve the Monge-Kantorovich problem in Equation 2 (Villani, 2021). The KR map is essentially a one-dimensional monotonically non-decreasing map, so it is easy to compute and estimate using an explicit construction formula. (Carlier et al., 2010) drew the link between the two maps by showing that the KR map is the limit of the Brenier map under a degenerated quadratic function.

Before introducing this linkage, let us define a twisted Euclidean cost function as the following:

$$c_{\beta,d}(x, y) = \tfrac{1}{2}(x - y)^\top A_{\beta,d}(x - y) = \tfrac{1}{2}\|A^{1/2}_{\beta,d}(x - y)\|_2^2 \tag{7}$$

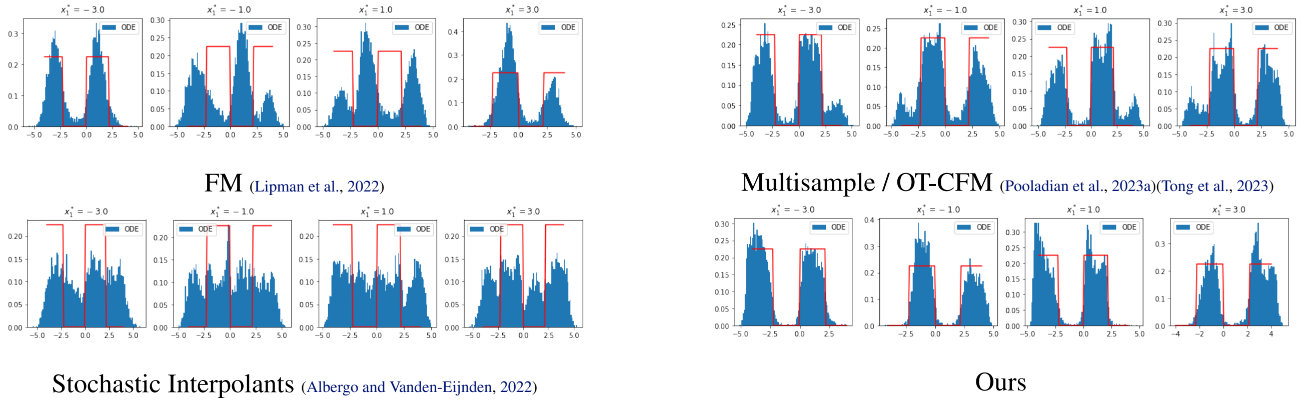
where $\beta \in (0, 1)$, take $\mathbf{1}_{d_1} := (1, \ldots, 1) \in \mathbb{R}^{d_1}$ and $\beta\mathbf{1}_{d_2} \in \mathbb{R}^{d_2}$ and define the matrix

$$A_{\beta,d} = \operatorname{diag}(\mathbf{1}_{d_1}, \beta\mathbf{1}_{d_2}) \in \mathbb{S}^d_{++}\,, \tag{8}$$

with $d = d_1 + d_2$. Although this illustration is in two dimensions, it can easily applied to higher dimensions.

Theorem 3.1. *(Carlier et al., 2010) Suppose ρ and μ have densities with respect to Lebesgue measure. Let $T_{\beta,d}$ define the optimal transport map with respect to the cost $c_{\beta,d}(x, y)$. Then the following convergence holds in $L^2(\rho)$: as $\beta \to 0$, $T_{\beta,d} \to T_{CB}$, where T_{CB} is given by Equation 3.*

Moreover, there is also a linkage between T_β and the KR map T_{KR} : Assumes that ρ is absolutely continuous with respect to Lebesgue measure and both ρ and μ have no atoms.



FM (Lipman et al., 2022)

Multisample / OT-CFM (Pooladian et al., 2023a)(Tong et al., 2023)

Stochastic Interpolants (Albergo and Vanden-Eijnden, 2022)

Ours

Figure 1: standard Gaussian to checkerboard; blue is the estimated conditional density; red is the true conditional density

Corollary 3.2 (Convergence to Knothe-Rosenblatt maps)**.** *(Carlier et al., 2010)(see Theorem 2.1) Let γ_β be an optimal plan with costs $c_{\beta,d}(x, y) = \Sigma_{i=1}^{d}\lambda_i(\beta)(x_i - y_i)^2$, for some weights $\lambda_i(\beta) > 0$. T_{KR} is the KR map between ρ,μ and γ_K the associated transport plan. Then $\gamma_\beta \to \gamma_K$ as $\beta \to 0$. if the plan γ_β is induced by $T_{\beta,d}$, then, as $\beta \to 0$, $T_{\beta,d} \overset{L^2(\rho)}{\to} T_{KR}$.*

These two results show that we can use the KR map to estimate the conditional Brenier map. By imposing the KR map structure through a twisted Euclidean cost function, the optimal transport map can converge to our desired conditional Brenier map (resp. Knothe-Rosenblatt map).

3.2. Drawing Conditional Samples

First, we train the JSFW following (Pooladian et al., 2023a) with the twisted Euclidean cost. To optimize Equation 6 in practice, we simply need to generate samples from the joint distribution q. When training from batches of samples $\{x^{(i)}\}_{i=1}^{n} \sim \rho$ and $\{y^{(i)}\}_{i=1}^{n} \sim \mu$, we can consider any doubly-stochastic matrix $\hat{q} \in \mathbb{R}_{+}^{n\times n}$ whose entries depend on the samples. Over data, the training objective then becomes

$$\min_{\theta} \sum_{i,j=1}^{n} \|v_\theta(t, (1-t)x^{(i)} + ty^{(j)}) - (y^{(j)} - x^{(i)})\|^2 \hat{q}(i, j) \tag{9}$$

where $t \sim \mathcal{U}([0, 1])$. As a special case, when $\hat{q}$ is chosen to solve Equation 2 with the empirical measures $\rho_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x^{(i)}}$ and $\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{y^{(i)}}$. In this setting, we can consider the optimal coupling matrix that solves (2) with cost c, denoted by $\hat{q}_c$. In this case, the coupling matrix is a permutation matrix, represented by $\sigma_c : \{1, \dots, n\} \to \{1, \dots, n\}$ (Peyré and Cuturi, 2019).

Algorithm 1 Conditional sampling with joint flow matching

```
**Input:** $\{x^{(i)}\}_{i=1}^{n} \sim \rho$ and $\{y^{(i)}\}_{i=1}^{n} \sim \mu$, $\beta$, $t$
**Step 1: train vector field** $v_\theta(t, x_t)$
**for** $k = 0, 1, 2, 3, \dots$ **do**
    source :$x_0$ = $\{x^{(i)}\}_{i=1}^{n}$ ; target : $x_1$ = $\{y^{(i)}\}_{i=1}^{n}$
    permutation matrix $\sigma_c$
    $\rightarrow$ *solve Monge-Kantorovich* ($x_0$,$x_1$,$\beta$)
    $x_1 = x_1^{\sigma_c}$
    $x_t = x_1 - x_0$
    train vector field $v_\theta(t, x_t)$
    loss $c_\theta$ = $\min_\theta \sum_{i=1}^{n} \|v_\theta(t, x_t) - (x_1^{\sigma_c(i)} - x_0)\|^2$
**end for**
**Step 2: integration using neural ODE**
**Input:** fix $y_1$ value and $y_2 \sim \rho_{y_2|y_1}$, $n$, $t$
initial condition $Y = [y_1, y_2]$ $\rightarrow$ *repeat n times*
sample = NeuralODE( trained $v_\theta(t, x_t)$, Y, t)
conditioned sample = sample[-1, : , dim:]
$\rightarrow$ *only need the last number of dims*
```

As a result, the training objective simplifies to

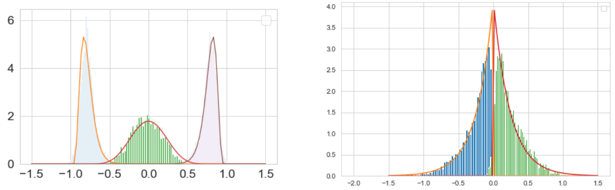
$$\min_{\theta} \sum_{i=1}^{n} \|v_\theta(t, (1-t)x^{(i)} + ty^{\sigma_c(i)}) - (y^{\sigma_c(i)} - x^{(i)})\|^2 \tag{10}$$

We use POT package (Flamary et al., 2021) to solve the Monge-Kantorovich problem in Equation 2 to get the permutation matrix. Finally, we use neural ODE (Chen et al., 2018) to numerically integrate the trained vector fields of generate the conditional samples. We summarize this procedure in Algorithm 1. See Appendix A for a detailed explanation. We provide an error analysis for sampling using 2-Wassersetin distance by following (Albergo and Vanden-Eijnden, 2022)'s proof strategy and expand it into the conditional sampling setting. See Appendix B for proof details.

3.3. Dimensional Reduction, Sparsity and Map Ordering

Estimating a transport map in high-dimensional settings is a challenging task. A common idea is to reduce the dimensions to a lower setting using PCA and VAE or project the probability measures onto a lower-dimensional subspace (Muzellec and Cuturi, 2019; Cuturi et al., 2023). Our approach reduces the dimensions by incorporating the twisted cost function to reduce the estimation complexity of the transport map, as this degenerate cost function penalizes the dimensions that the input data is conditioned on.

Moreover, our approach adds more sparsity to the existing sparsity presented in the lower triangular transport map structure to promote faster training and sampling efficiency. Whereas the conditional Brenier map is a dense lower triangular map. (Spantini et al., 2018) noted that the KR map can be considered as imposing the sparsest structure while still preserving the coupling. A key advantage of using this type of map for conditional sampling is that it is anisotropic dependent on the input data dimensions. The KR map can capture the conditional distribution without the need for each component of the map to depend on the entire input data dimensions.



$y = \tanh(x+z),\ z \sim \mathcal{N}(0, 0.05)$ $y = z\tanh(x),\ z \sim \Gamma(1, 0.3)$

Figure 2: The line plots are the true conditional density. The histograms are the approximated distribution using our approach

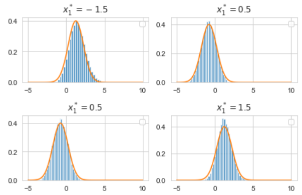


Figure 3: Normal order

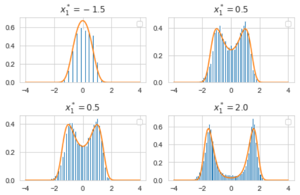


Figure 4: Reverse order

Furthermore, another approach to increase the sparsity of the KR map further is to change its ordering to find an optimal permutation. This is an NP-hard problem and relies on heuristics but prior research has shown using methods such as reversed Cholesky (Saad, 2003) or min-fill and min-degree methods (Koller and Friedman, 2009) can tackle this issue. (Morrison et al., 2017) also explored the linkage between the sparsity of the transport map and the sparsity of the probabilistic graphical model where reducing independent edges creates more sparsity in the transport map. Our method does not explore this direction, and the KR map follows a monotonically non-decreasing order.

Table 1: Image In-painting / 2D Star results

	CIFAR10					$\mathcal{N} \sim$ Star
FID	@ NFE=2	4	6	8	10	W_2^2
Ours	44.90	21.69	19.94	19.75	**18.72**	**0.096**
Multisample-FM OT-FM (Pooladian et al., 2023a) (Tong et al., 2023)	47.87	23.72	22.09	21.13	20.58	0.127
Independent-FM (Tong et al., 2023)	44.32	23.12	20.73	20.78	20.14	0.543
Stochastic interpolants (Albergo and Vanden-Eijnden, 2022)	187.67	39.04	30.47	26.07	20.58	0.146

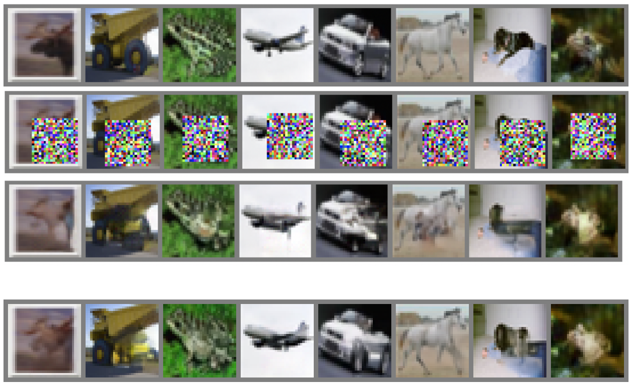


Figure 5: CIFAR10 image in-painting. **order:** Original image, Masked image, Ours at 4k steps,NFE=10, Ours at 100k steps, NFE=10

4. Related work

Conditional sampling with measure transport Previous works explored using a neural network-based approach to approximate the transport map as a convex function for unconditional tasks like color transfer (Makkuva et al., 2020; Korotin et al., 2022; Uscidda and Cuturi, 2023). Our work is interested in drawing the linkage between optimal transport, generative model and conditional sampling to solve conditional tasks such as inverse problems like image in-painting and super resolution. In particular, we are interested in tackling this problem from the measure transport approach (Marzouk et al., 2016). Along this line, two recent works are closely related to our approach: (Kovachki et al., 2020) introduced Monotone-GAN to perform conditional sampling through a trained generator as the approximated transport map, where the estimated map is also a lower triangular shape with an imposed monotonically increasing constraint and (Shi et al., 2022) uses a Schrodinger bridge based diffusion model to approximate the conditional distribution.

Figure 7: Conditioning with our SR method

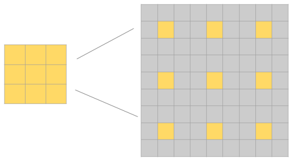


Table 2: CelebA SR 4X

NFE =20	FID ↓	PSNR ↑	SSIM ↑
Ours	**25.23**	**23.84**	**0.793**
CDSB	57.22	19.72	0.504
CSGM	92.02	19.52	0.471
CSGM-C	44.44	20.44	0.566
CDSB-C	28.41	21.11	0.614

Table 3: ImageNet128 SR 2X

$\sigma = 0.05$	FID ↓	PSNR ↑	SSIM ↑
Ours	20.45	27.03	**0.885**
ΠGDM	4.38	32.07	0.831
OT-ODE	4.61	**32.59**	0.862
RED-Diff	10.54	31.82	0.852
D-flow	**4.26**	32.35	0.858

Figure 8: **Left:** this bigger matrix is our input source image where conditioned pixels are in yellow, also added noise in gray; **Middle:** compares results reported from (Shi et al., 2022); **Right:** compares results reported from ΠGDM (Song et al., 2022),OT-ODE(Pokle et al., 2023),RED-Diff (Mardani et al., 2023) and D-flow (Ben-Hamu et al., 2024) see related works

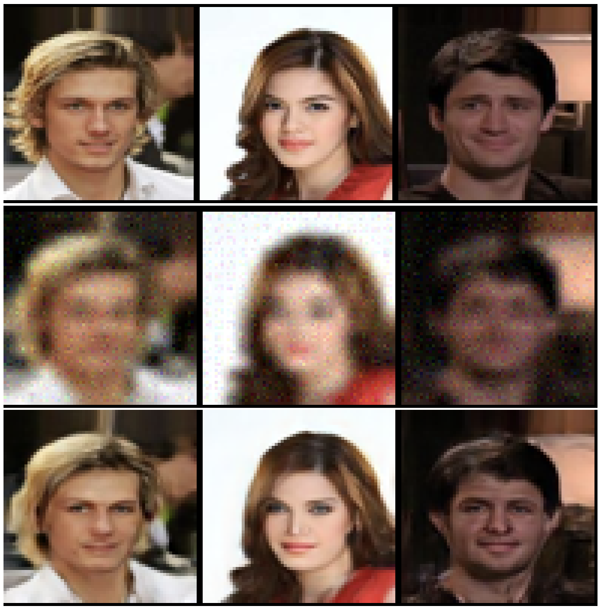


Figure 6: CelebA64 SR 4X from 16×16 to 64×64 **order:** Ground truth, Input image, Output image

Unlike these works, our work uses joint flow matching, which enjoys faster training time, efficient sampling, and less computational cost. Moreover, we incorporate a twisted Euclidean cost to impose a sparse lower triangular transport map called KR rearrangement to solve the optimal coupling matrix. Using a diverse cost function to estimate transport maps has been a recent development where (Klein et al., 2023; Pooladian et al., 2023b; Neklyudov et al., 2023) have explored using diverse costs from a computational lens. Our method ventures in this direction for conditional sampling. Previously, it was a standard to use the squared-Euclidean cost to estimate the transport map and is widely applied in economics, computational biology, and computer graphics (Bunne et al., 2022; Solomon et al., 2015). Statistical estimation of these maps was made rigorous in (Hütter and Rigollet, 2021), which was followed swiftly by (Deb et al., 2021; Manole et al., 2021; Pooladian and Niles-Weed, 2021).

Other diffusion/flow-based approaches Other than using measure transport to estimate the true posterior distribution, previous works also explored using pre-trained flow matching models (Pokle et al., 2023) to solve inverse problems, using variational inference with diffusion models (Mardani et al., 2023), incorporating the conditions as prior (Saharia et al., 2022) or differentiating through the generative process of a flow or diffusion model by solving an optimization problem (Ben-Hamu et al., 2024).

5. Experiments

In this section, we validate our approach through a series of numerical experiments from lower dimensional data distribution to solving linear inverse problems such as image in-painting and super resolution. See Appendix C for more.

5.1. 2D Synthetic data

Conditional tanh functions We test the approach on two-dimensional non-linear and non-Gaussian distribution examples from (Kovachki et al., 2020) to illustrate the conditional sampling quality. We draw $x \sim \mathcal{U}(-3, 3)$ and keep $x \in [-1.1, 0, 1.1]$ fixed for the experiments. These equations have closed-form solutions for the true posteriors $\mathcal{P}(y|x)$, which are the y values. The source distribution is defined as a joint distribution of $(\mathcal{U}[-3, 3], \mathcal{N}[0, 1])$. Figure 2 compares the approximated conditional sampling distribution using our transport map in bars with the true data distribution in lines.

Reverse ordering This experiment aims to test the robustness of our method by reversing the variable order of a 2D half-moon-shape distribution. Changing the variable orders adds more complexity to our estimated transport

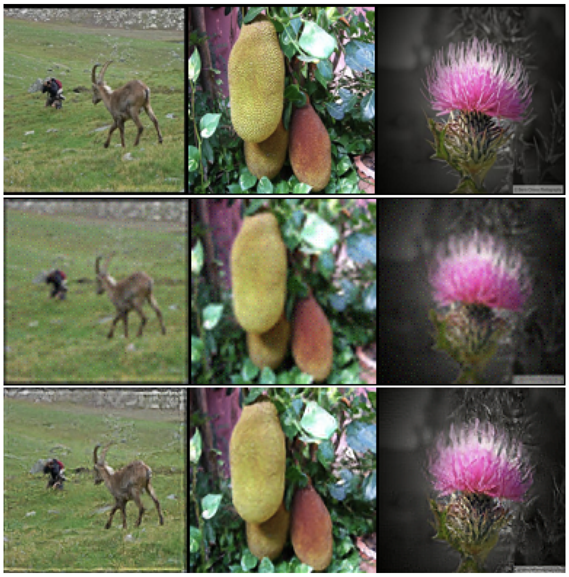


Figure 9: ImageNet128 SR 2X from 64 × 64 to 128 × 128. **order:** Ground truth, Input image, Output image

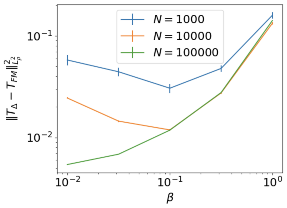


Figure 10: conditional Brenier map estimation error

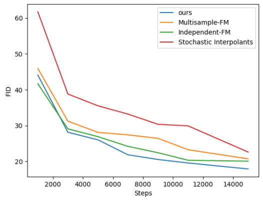


Figure 11: FID vs. Steps on Cifar10

map to approximate the conditional distribution. We use $x_1 \sim \mathcal{N}(0, 1)$ and $x_2 = x_1^2 - 1$ to create the half-moon shape. Instead of the usual ordering to assemble the joint distribution, we now reverse the orders and define the source distribution as $\rho = [x_2, x_1]$ and the conditional distribution $\mu(x_1|x_2)$ is defined as $\mathcal{N}(x_1^2 - 1, 1)$. Figure 3 and Figure 4 shows that our approximated transport map can capture the conditional distribution for both the normal and reverse order distributions.

Model Comparison We first qualitatively compared conditional sampling quality among various flow matching approaches in Figure 1 where the source is a standard Gaussian and the target is a 2D checkerboard distribution. Our approach was the closest to generating samples within the true conditional density in red. All the models were trained for 2k iterations. Moreover, we quantitatively compared these methods in terms of fit through 2-Wasserstein distances in Table 1 which use standard Gaussian as source distribution, and the star distribution as target and ours has the lowest W_2^2 distance.

5.2. Conditional Brenier Map Estimation

In this experiment, we first test our approach on a 4-dimensional Gaussian distribution as a target and compare it with the true conditional density. Furthermore, we compute the approximation error between the true conditional Brenier map and our estimated transport map using JSFM in L_2. We fix one dimension and infer the rest 3-dimensional θ. In Figure 10, we measure the approximation errors between the true conditional Brenier map and our estimated map and its relationship with the change of the small β value used in our cost function. Figure 10 indicates the error increases as β grows since a larger β value has less penalize power on the conditioned dimensions and does not align with Corollary(3.2) for $T_\beta \to T_{\text{KR}}$ convergence.

5.3. Image In-painting

On the CIFAR10 dataset, we masked a randomly placed 20 by 20 square patch on each image denoted as y and the rest of the image denoted as x, where the final result is a sample from $\mu(y|x)$. We add standard Gaussian noise onto the masked patch y to create the joint distribution as input data. To follow Equation 7, the small beta values are applied to unmasked pixels, and the input source image x_0 is flattened as a requisite for the KR map structure. After solving for optimal coupling, we assign the indices with the largest value to the target sample x_1, indicating moving the largest possible optimal "mass" to the destination. Furthermore, we used the TorchCFM package (Tong et al., 2023) for comparison with other variations of flow matching.

Conditional Sampling quality We investigated the sample quality using FID score among various conditional sampling approaches. Table 1 shows our method achieves the lowest FID as NFE increases. At a given NFE, our approach almost always has the lowest FID.

Training Efficiency Figure 5 shows the results from our method after 100k steps. We also noticed that our approach was able to generate the high-level shape of the objects at as low as 4k iterations and 10 NFEs. Moreover, Figure 11 shows our approach in blue line has the lowest FID score as the training step progresses. (Pooladian et al., 2023a) proved that using joint flow matching can reduce the gradient variance during training, thus leading to faster training.

5.4. Super resolution

We also tested on super resolution using the CelebA dataset and ImageNet-128 with 1k classes. For CelebA, we experimented on SR 4X from 16 × 16 low resolution to 64 × 64. The input images use bi-cubic interpolation to the downsampled images. For super resolution, we add Gaussian noise to the input image and conditioned on the downsampled low-resolution image, for example, the 16 × 16 image for CelebA. Note that Gaussian noise was not added to down-sampled pixel locations for the input image as illustrated in Figure 7. In Table 2, we compare the results with different variations from (Shi et al., 2022) where CDSB-C is a Schrodinger bridge-based diffusion model, CSGM-C is a conditional score-based model and both added Gaussian noise. Furthermore, we experimented on ImageNet-128 for SR 2X, trained for 500K steps, and compared with the most recent conditional generation state-of-the-art approaches using flow matching and diffusion models shown in Table 3. While the PSNR and SSIM are comparable with the state-of-the-art methods, we noticed the gap in the FID score. A limitation of this paper is the need to train more steps to further reduce the FID score for much higher dimensional data. In contrast, our result from 2 trained for 300K outperformed other methods trained for 500k.

6. Conclusions

In this work, we strengthened the connections between applications in optimal transport and joint flow matching. By choosing a particular cost function, we exploited the underlying limiting structure of the learned vector field, which provides an efficient method for conditional sampling. The limitation of this work is to develop tighter bound statistical guaranties. Future works could include other easy-to-implement cost functions that lead to desired behavior in the limiting flows.

References

Youssef Marzouk, Tarek Moselhy, Matthew Parno, and Alessio Spantini. An introduction to sampling via measure transport. *arXiv preprint arXiv:1602.05023*, 2016.

Nikola Kovachki, Ricardo Baptista, Bamdad Hosseini, and Youssef Marzouk. Conditional sampling with monotone gans. *arXiv preprint arXiv:2006.06755*, 2020.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky Chen. Multisample flow matching: Straightening flows with minibatch couplings. *arXiv preprint arXiv:2304.14772*, 2023a.

Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio. From knothe's transport to brenier's map and a continuation method for optimal transport. *SIAM Journal on Mathematical Analysis*, 41(6):2554–2576, 2010.

Filippo Santambrogio. Optimal transport for applied mathematicians. 2015.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

L. Kantorovitch. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Zhaoyue Chen, Mokhwa Lee, and Yifan Sun. Continuous time frank-wolfe does not zig-zag, but multistep methods do not accelerate, 2022.

Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian Fatras, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023.

Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *The Journal of Machine Learning Research*, 22(1):3571–3578, 2021.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31, 2018.

Boris Muzellec and Marco Cuturi. Subspace detours: Building transport plans that are optimal on subspace projections. *Advances in Neural Information Processing Systems*, 32, 2019.

Marco Cuturi, Michal Klein, and Pierre Ablin. Monge, bregman and occam: Interpretable optimal transport in high-dimensions with feature-sparse maps. *arXiv preprint arXiv:2302.04065*, 2023.

Alessio Spantini, Daniele Bigoni, and Youssef Marzouk. Inference via low-dimensional couplings. *Journal of Machine Learning Research*, 19(66):1–71, 2018.

Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

Rebecca Morrison, Ricardo Baptista, and Youssef Marzouk. Beyond normality: Learning sparse probabilistic graphical models in the non-gaussian setting. *Advances in neural information processing systems*, 30, 2017.

Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Conditional simulation using diffusion schrödinger bridges. In *Uncertainty in Artificial Intelligence*, pages 1792–1802. PMLR, 2022.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2022.

Ashwini Pokle, Matthew J Muckley, Ricky TQ Chen, and Brian Karrer. Training-free linear image inversion via flows. *arXiv preprint arXiv:2310.04432*, 2023.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. *arXiv preprint arXiv:2305.04391*, 2023.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. *arXiv preprint arXiv:2402.14017*, 2024.

Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pages 6672–6681. PMLR, 2020.

Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Neural optimal transport. *arXiv preprint arXiv:2201.12220*, 2022.

Théo Uscidda and Marco Cuturi. The monge gap: A regularizer to learn all transport maps. *arXiv preprint arXiv:2302.04953*, 2023.

Michal Klein, Aram-Alexandre Pooladian, Pierre Ablin, Eugène Ndiaye, Jonathan Niles-Weed, and Marco Cuturi. Learning costs for structured monge displacements. *arXiv preprint arXiv:2306.11895*, 2023.

Aram-Alexandre Pooladian, Carles Domingo-Enrich, Ricky TQ Chen, and Brandon Amos. Neural optimal transport with lagrangian costs. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023b.

Kirill Neklyudov, Rob Brekelmans, Alexander Tong, Lazar Atanackovic, Qiang Liu, and Alireza Makhzani. A computational framework for solving wasserstein lagrangian flows. *arXiv preprint arXiv:2310.10649*, 2023.

Charlotte Bunne, Andreas Krause, and Marco Cuturi. Supervised training of conditional monge maps. *Advances in Neural Information Processing Systems*, 35:6859–6872, 2022.

Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4):1–11, 2015.

Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. 2021.

Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. *Advances in Neural Information Processing Systems*, 34:29736–29753, 2021.

Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *arXiv preprint arXiv:2107.12364*, 2021.

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4): 4713–4726, 2022.

Ricky T. Q. Chen. torchdiffeq, 2018. URL https://github.com/rtqichen/torchdiffeq.

A. Details on the algorithm

We give an example on generating conditioned samples using a 2D standard Gaussian distribution as source ρ and another Gaussian distribution as target μ. First, we draw mini-batch samples from both distributions. Second, we solve the Monge-Kantorovich problem to find the optimal coupling between the source and target mini-batches samples. This allows us to permute the ordering of the samples to have an optimal matching between the two distributions. We then define the conditional path as a straight line and interpolate it as $x_t = (1-t)x_1 + tx_2$. Third, we train the vector fields $v_\theta(t, x_t)$ using neural nets and use the neural ODE method through torchdiffeq package (Chen, 2018) for integration to get the conditioned samples. For this step, we first create a joint distribution as the initial input data. We start by choosing and then fixing the value of the first dimension where $x_1 \in \mathbb{R}^{d_1}$, then randomly draw a sample from the standard Gaussian distribution as the second dimension $x_2 \sim \rho$. For more complicated distributions, we need to draw x_2 conditioned on this fixed x_1 as $x_2 \sim \rho(x_2|x_1)$ since we can not draw the samples independently. In some cases, we can also get closed-form solutions. Next, we stack these two dimensions to form $X := [x_1, x_2] \sim \rho$ as the joint source distribution. Finally, we can use the neural ODE to generate conditional samples. The conditioned sample is the last dimension in this 2D case, and it represents a sample drawn from our *approximated* distribution $\mu_{2|1}(\cdot|x_1)$.

B. Error analysis on conditional sampling

We use 2-Wassersetin distance as the metric to analyze the conditional sampling error bound between the estimated conditional pdf $\hat{u}_t(\cdot|z)$ and $u_t(\cdot|z)$ conditioned on fixed variable z.

$$W_2^2(\hat{u}_t(\cdot|z), u_t(\cdot|z)) \leq \exp^{1+2\hat{K}} Q(\hat{v}_t(\cdot|z)), \tag{11}$$

where $\hat{K}$ is a Lipschitz constant and $Q(\hat{v}_t)$ is defined as

$$\text{Q}(_t(\cdot|z)) = \int_0^1 \int_{\mathbb{R}d} \|v_t(X_t(\cdot|z)) - v_t(\hat{X}_t(\cdot|z))\|^2 \pi_0(dX_t, d\hat{X})dt \tag{12}$$

Proof: first, by definition we have

$$W_2^2(\hat{u}_t(\cdot|z), u_t(\cdot|z)) := \inf \int_{\mathbb{R}d} \|X_t(x|z) - \hat{X}_t(x|z)\|^2 \pi_0(dX_t, d\hat{X}) \leq \int_{\mathbb{R}d} \|X_t(x|z) - \hat{X}_t(x|z)\|^2 \pi_0(dX_t, d\hat{X})$$

Then define $\dot{X}_t(x|z) = v_t(X_t(x|z))$ and $\dot{\hat{X}}_t(x|z) = v_t(\hat{X}_t(x|z))$ and

$$Y_t = \int_{\mathbb{R}d} \|X_t(x|z) - \hat{X}_t(x|z)\|^2 \pi_0(dX_t, d\hat{X}) \tag{13}$$

Then differentiate on both sides, we get

$$\begin{aligned}\dot{Y}_t &= 2\int_{\mathbb{R}d} \langle\, X_t(x|z) - \hat{X}_t(x|z), v_t(X_t(x|z) - v_t(\hat{X}_t(x|z)\rangle \pi_0(dX_t, d\hat{X}) \\ &= 2\int_{\mathbb{R}d} \langle\, X_t(x|z) - \hat{X}_t(x|z), v_t(X_t(x|z) - v_t(\hat{X}_t(x|z) + v_t(\hat{X}_t(x|z) - v_t(\hat{X}_t(x|z)\rangle \pi_0(dX_t, d\hat{X})\end{aligned}$$

Then using $2\langle a, b\rangle \leq \|a\|^2 + \|b\|^2$

$$\begin{aligned}&2\langle\, X_t(x|z) - \hat{X}_t(x|z), v_t(X_t(x|z) - v_t(\hat{X}_t(x|z) \\ &\leq \|X_t(x|z) - \hat{X}_t(x|z)\|^2 + \|v_t(\hat{X}_t(x|z) - v_t(\hat{X}_t(x|z)\|^2\end{aligned}$$

Moreover because $\hat{X}_t(x|z)$ and $X_t(x|z)$ satisfies Lipschitz continuous, we also have the following inequality:

$$\begin{aligned}&2\langle\, X_t(x|z) - \hat{X}_t(x|z), v_t(X_t(x|z) - v_t(\hat{X}_t(x|z)\rangle \\ &\leq 2\hat{K}_t \|\hat{X}_t(x|z) - X_t(x|z)\|^2\end{aligned}$$

Putting the above two inequalities together, we have

$$\dot{Y}_t \leq (1 + 2\hat{K})Y_t + \int_{\mathbb{R}d} \|v_t(X_t(x|z)) - v_t(\hat{X}_t(x|z))\|^2 \pi_0(dX_t, d\hat{X})$$

Finally, by Gronwall's inequality and since $Y_0 = 0$, and integrating $\dot{Y}_t$, we have the following

$$\int \dot{Y}_t \pi(dz) \leq \exp^{1+2\hat{K}} \int_0^1 \int_{\mathbb{R}d} \|v_t(X_t(x|z)) - v_t(\hat{X}_t(x|z))\|^2 \pi_0(dX_t, d\hat{X})dt$$

Next,define $Q(\hat{v}_t)$ as

$$Q(\hat{v}_t) = \int_0^1 \int_{\mathbb{R}d} \|v_t(X_t(x|z)) - v_t(\hat{X}_t(x|z))\|^2 \pi_0(dX_t, d\hat{X})dt$$

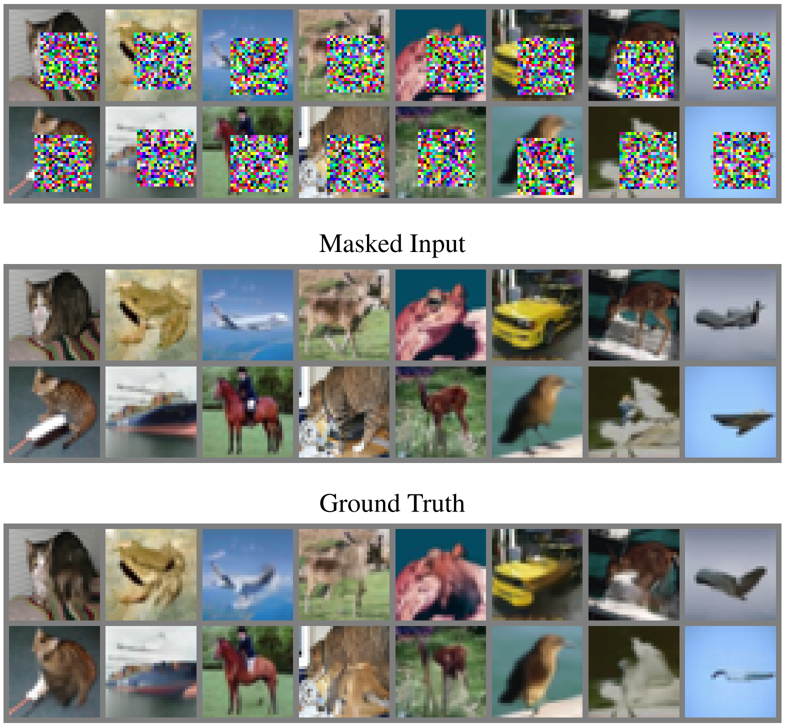
Then by (Albergo and Vanden-Eijnden, 2022) Equation 27, we also have

$$W_2^2(\hat{u}_t(\cdot|z), u_t(\cdot|z)) \leq \int \dot{Y}_t \pi(dz)$$

Thus, finally we obtain the error bound as

$$W_2^2(\hat{u}_t(\cdot|z), u_t(\cdot|z)) \leq \exp^{1+2\hat{K}} Q(\hat{u}_t)$$

C. Additional Results



Masked Input

Ground Truth

Output Image

Figure 12: Image Inpainting results from CIFAR10 at 100k steps

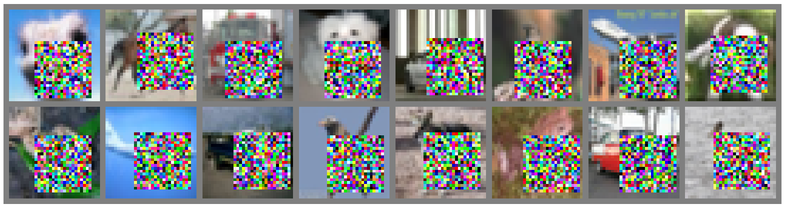
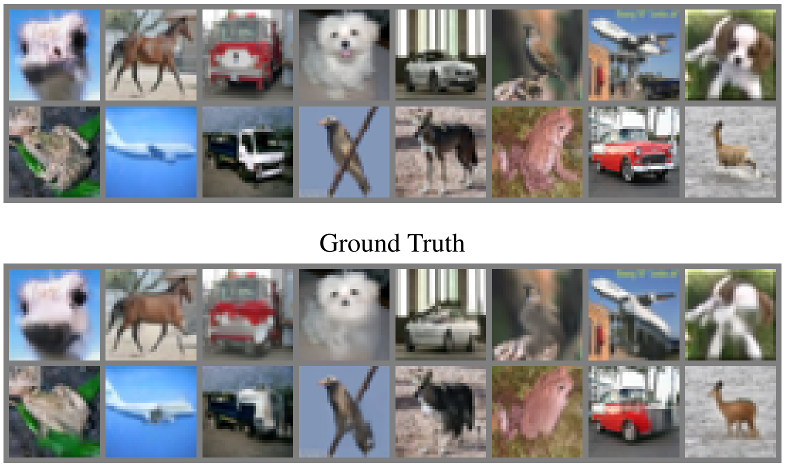


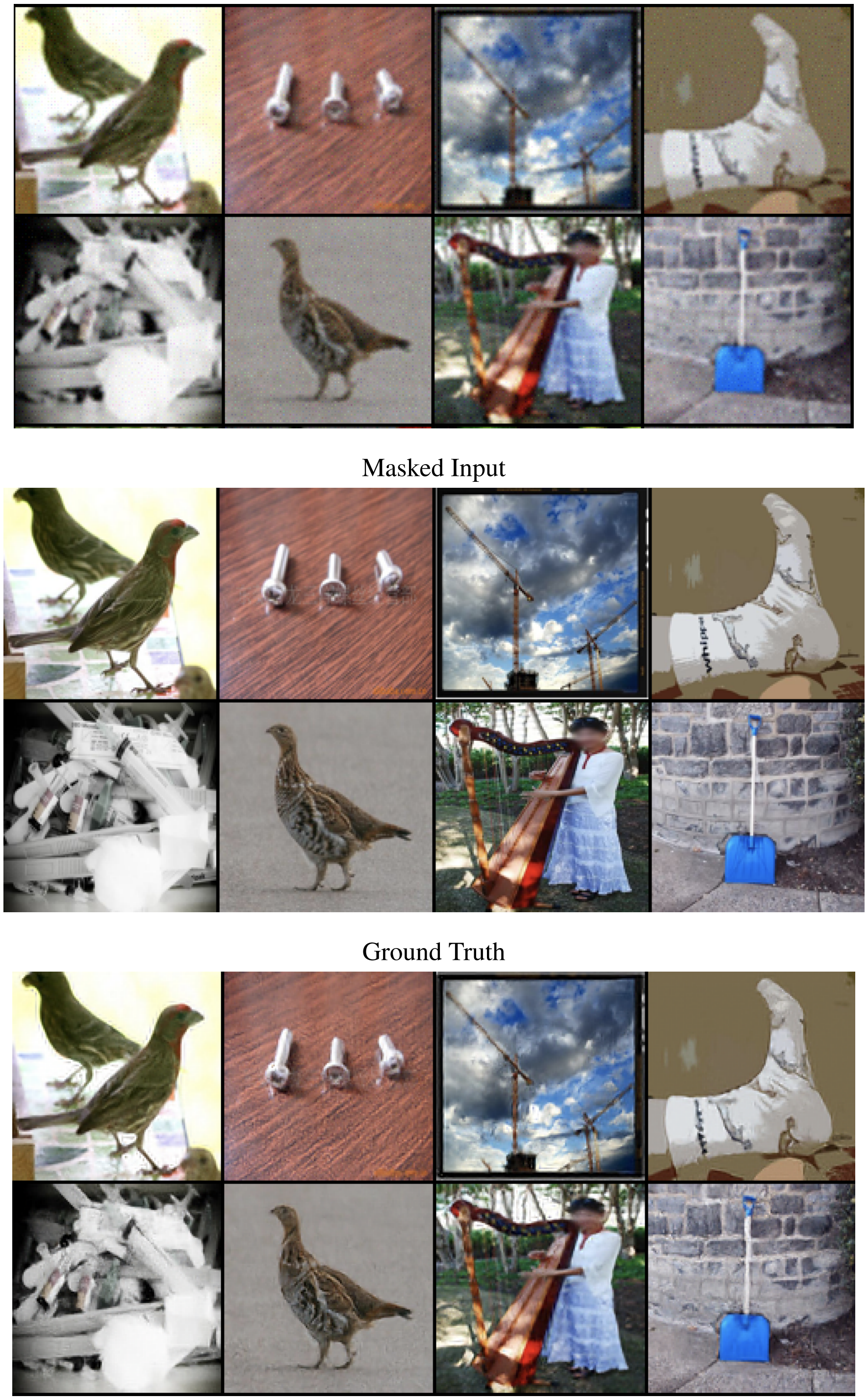
Figure 13: Masked Input



Ground Truth

Output Image

Figure 14: Image Inpainting results from CIFAR10 at 100k steps

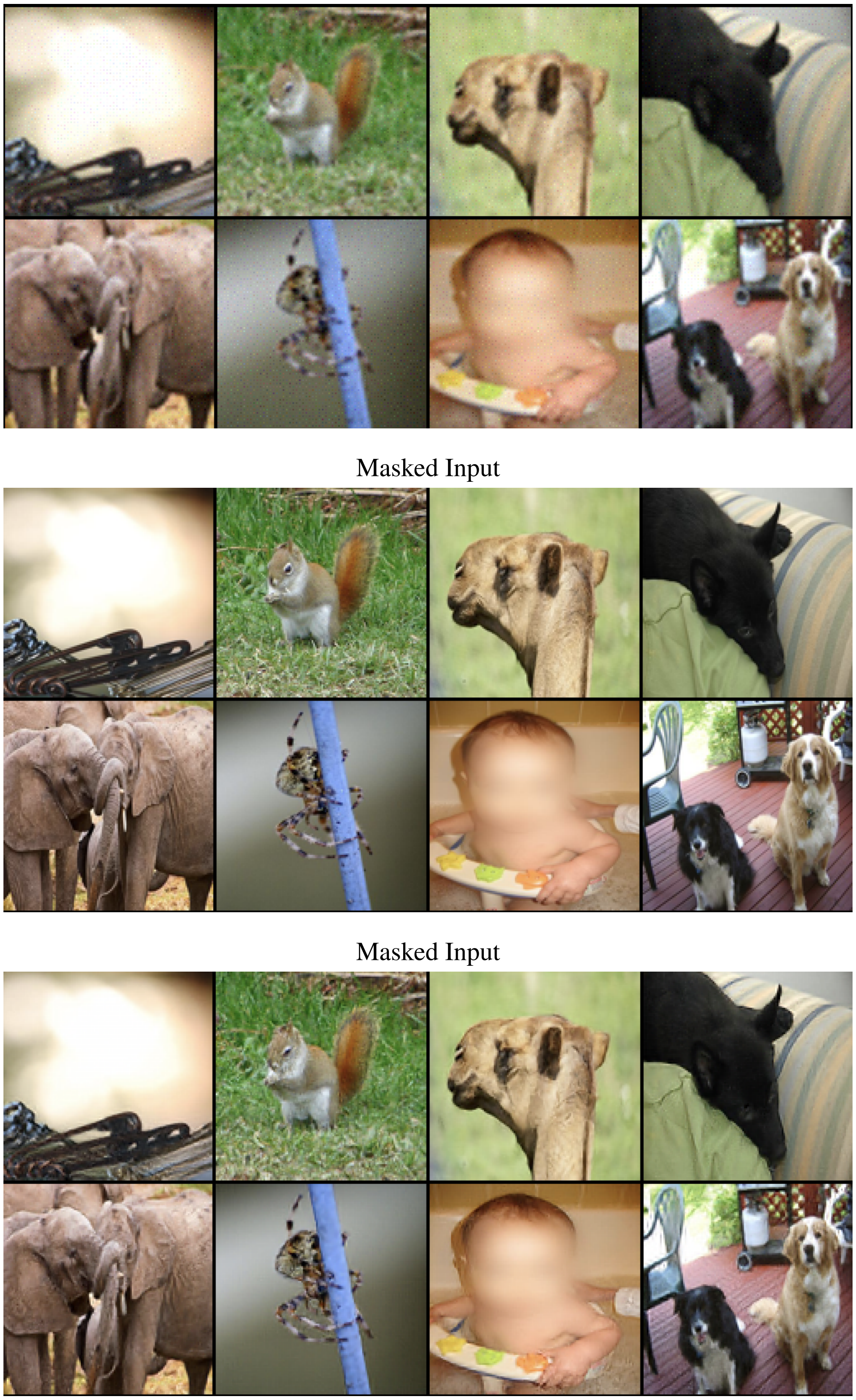


Masked Input

Ground Truth

Output Image

Figure 15: Super resolution 2X 64 to 128 on ImageNet-1k at 500k steps

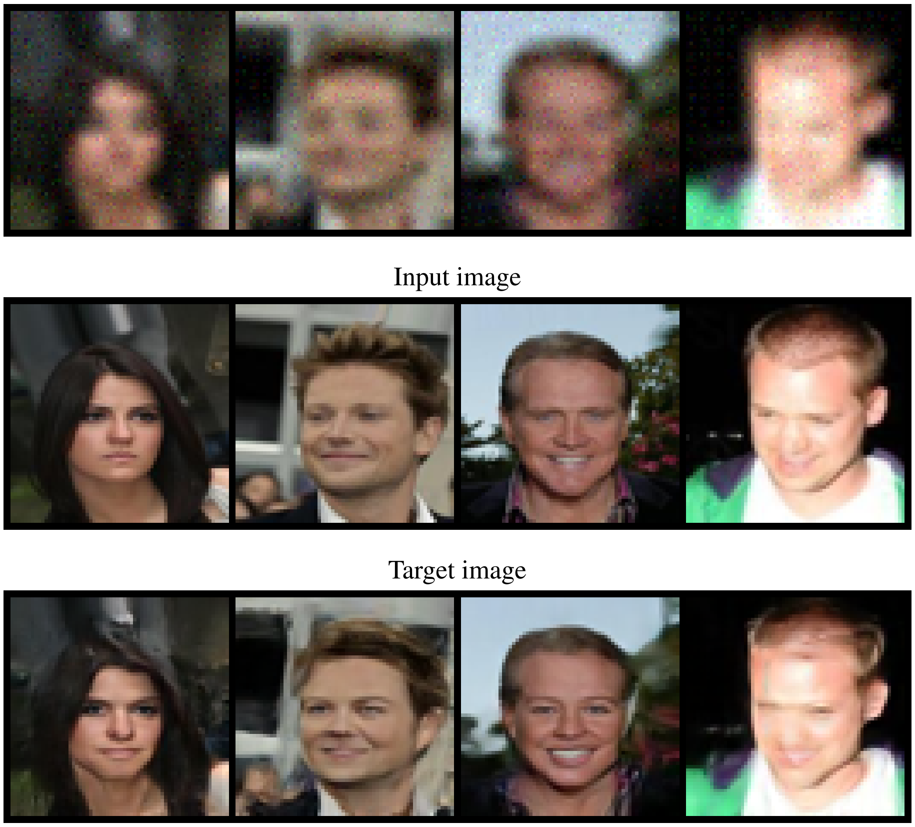


Masked Input

Masked Input

Output Image

Figure 16: Super resolution 2X 64 to 128 on ImageNet-1k at 500k steps

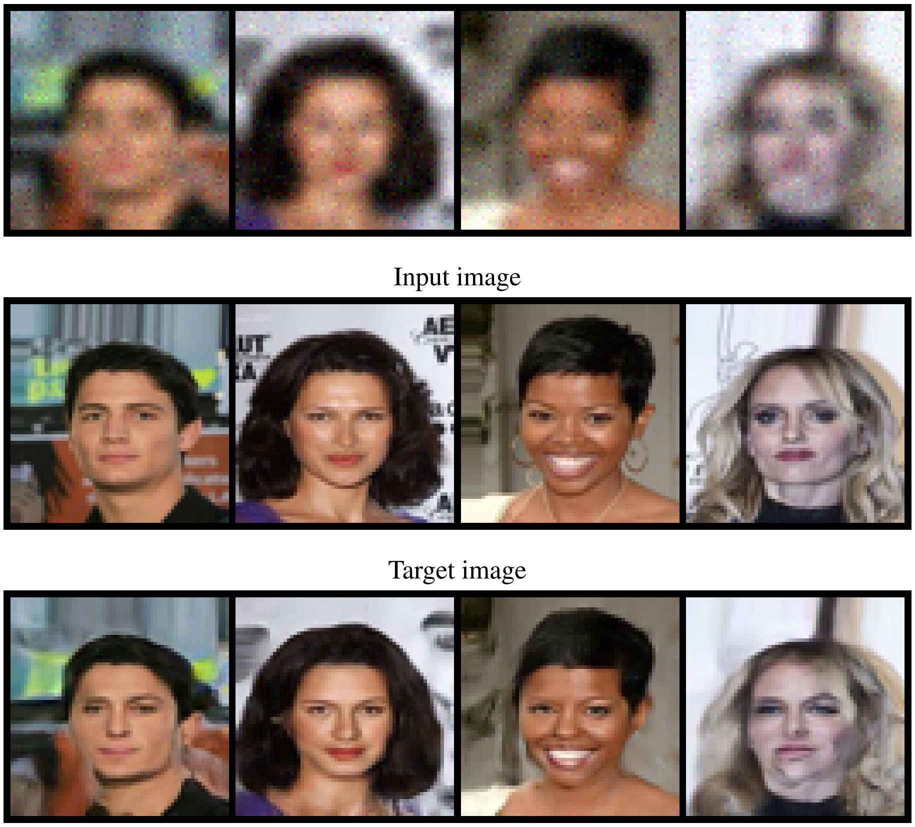


Input image

Target image

Output image

Figure 17: Super resolution 4X 16 to 64 on CelebA at 300k steps



Input image

Target image

Output image

Figure 18: Super resolution 4X 16 to 64 on CelebA at 300k steps