Nonparametric Generative Modeling with Conditional Sliced-Wasserstein Flows

Chao Du, Tianbo Li, Tianyu Pang, Shuicheng Yan, Min Lin

Sea AI Lab

TL;DR: A nonparametric conditional generative model (without NN & SGD training) achieves promising results.

Introduction

Nonparametric methods enjoy infinite capacity and flexibility but have been less explored in generative modeling. In these work, we

- Reveal the **conditional modeling capabilities** of SWF
- Introduce inductive biases for image tasks into SWF

Main takeaways:

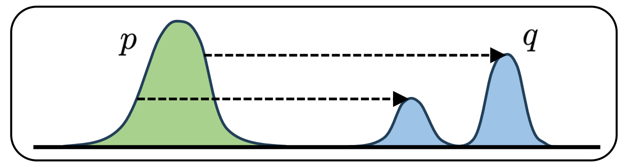
- The first nonparametric conditional generative model
- Achieve comparable performance with parametric generative models

Paper

Background

1D Wasserstein distance (closed form):

$$W_2^2(p,q) = \int_0^1 |F_p^{-1}(\tau) - F_q^{-1}(\tau)|^2 \mathrm{d}\tau$$

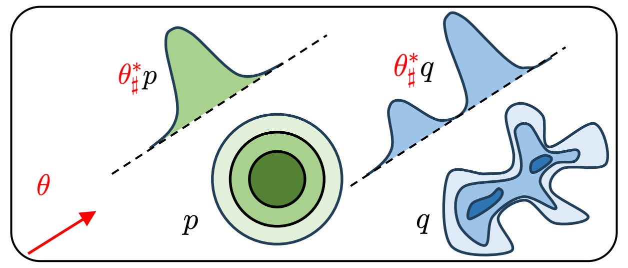


Inverse cumulative distribution functions (CDF) Associated transport map: $T = F_q^{-1} \circ F_p$

Sliced-Wasserstein distance:

Based on projections: $\theta^*(x) \triangleq \langle \theta, x \rangle$

$$SW_2^2(p,q) \triangleq \int_{\mathbb{S}^{d-1}} W_2^2(\theta_\sharp^* p, \theta_\sharp^* q) \mathrm{d}\theta$$



SWF: Gradient flow in the Wasserstein space minimizing SW_2

$$\min_{p \in \mathcal{P}_2(\mathbb{R}^d)} SW_2^2(p,q) \quad \Rightarrow \quad \frac{\partial p_t(x)}{\partial t} + \nabla \cdot (p_t(x) v_t(x)) = 0$$

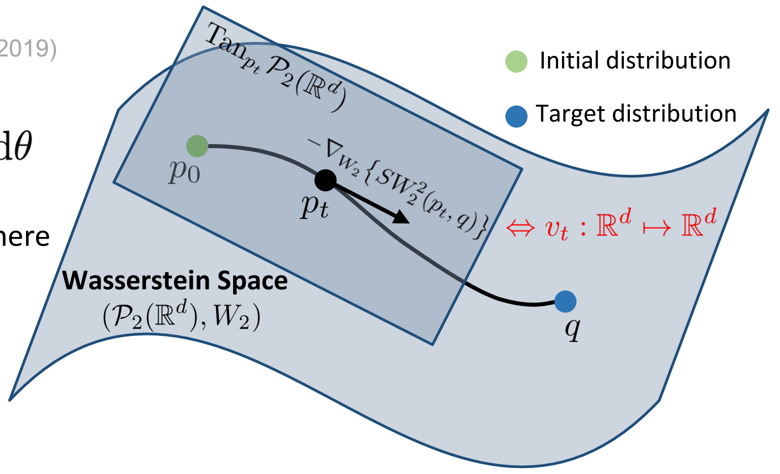
Velocity field: (Bonnotte 2013; Liutkus et al., 2019)

$$v_t(x) \triangleq -\int_{\mathbb{S}^{d-1}} \psi'_{t,\theta}(\theta^\top x) \cdot \theta \, \mathrm{d}\theta$$

Monte Carlo estimation over unit sphere

$$\psi'_{t,\theta}(z) = z - F_{\theta_\sharp^* q}^{-1} \circ F_{\theta_\sharp^* p_t}(z)$$

one-dimensional CDF estimations



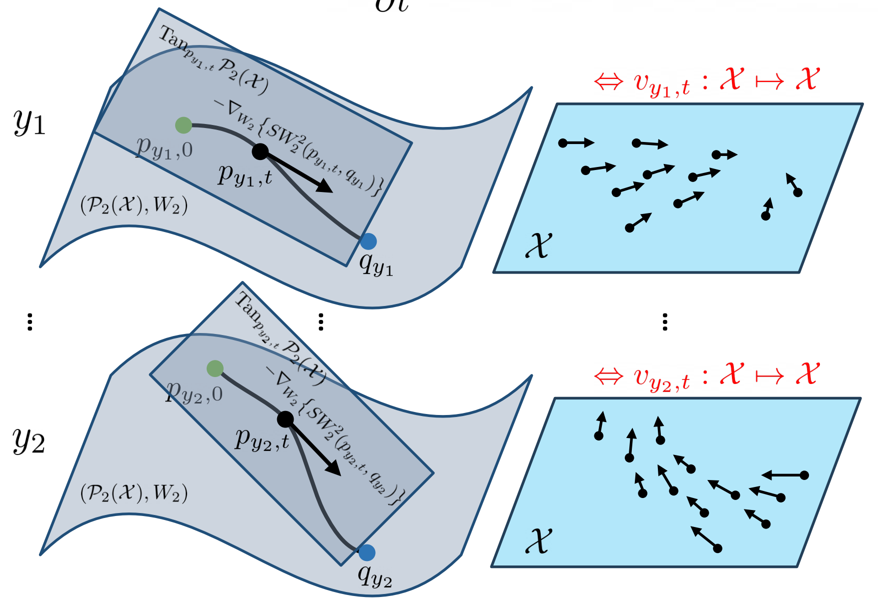
Conditional SWF

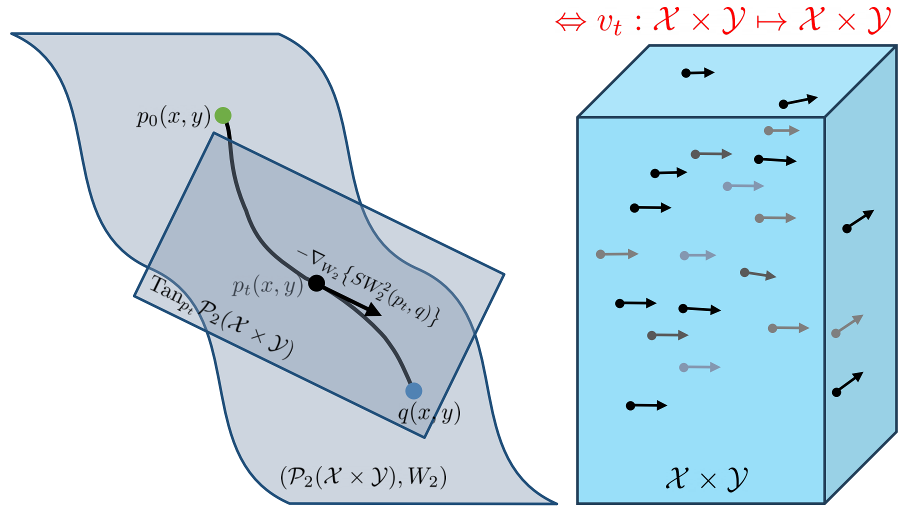
Ideally: Collection of SWFs $\min_{p_y \in \mathcal{P}_2(\mathcal{X})} SW_2^2(p_y, q_y), \forall y \in \mathcal{Y}$

$\Rightarrow$ $(p_{y,t})_{t \geq 0}$ solves $\frac{\partial p_{y,t}(x)}{\partial t} + \nabla \cdot (p_{y,t}(x) v_{y,t}(x)) = 0$

This work: SWF in the joint space $\min_{p \in \mathcal{P}_2(\mathcal{X} \times \mathcal{Y})} SW_2^2(p,q)$

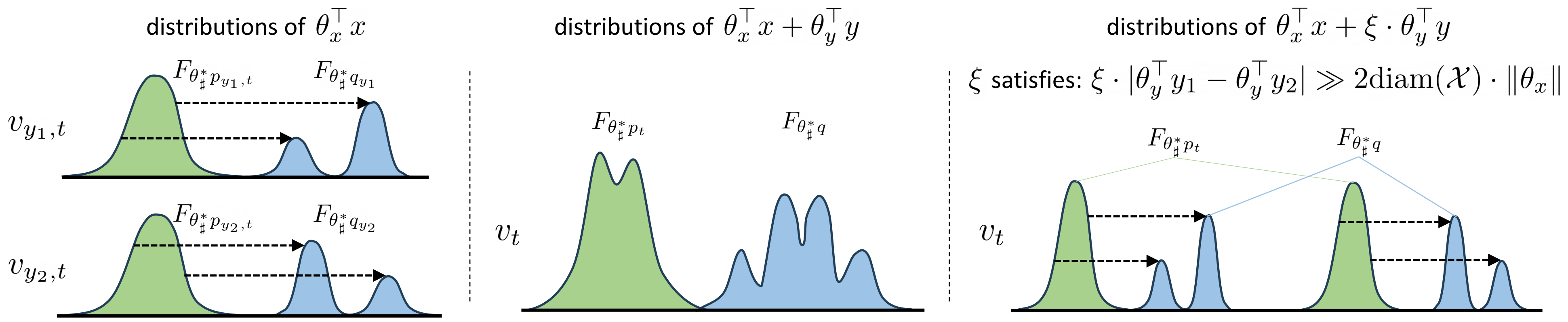
$\Rightarrow$ $(p_t)_{t \geq 0}$ solves $\frac{\partial p_t(x,y)}{\partial t} + \nabla \cdot (p_t(x,y) v_t(x,y)) = 0$





Observation: (under certain conditions) the velocities coincide! $v_t^{\mathcal{X}}(x,y) \approx v_{y,t}(x)$ $\quad v_t^{\mathcal{Y}}(x,y) \approx 0$

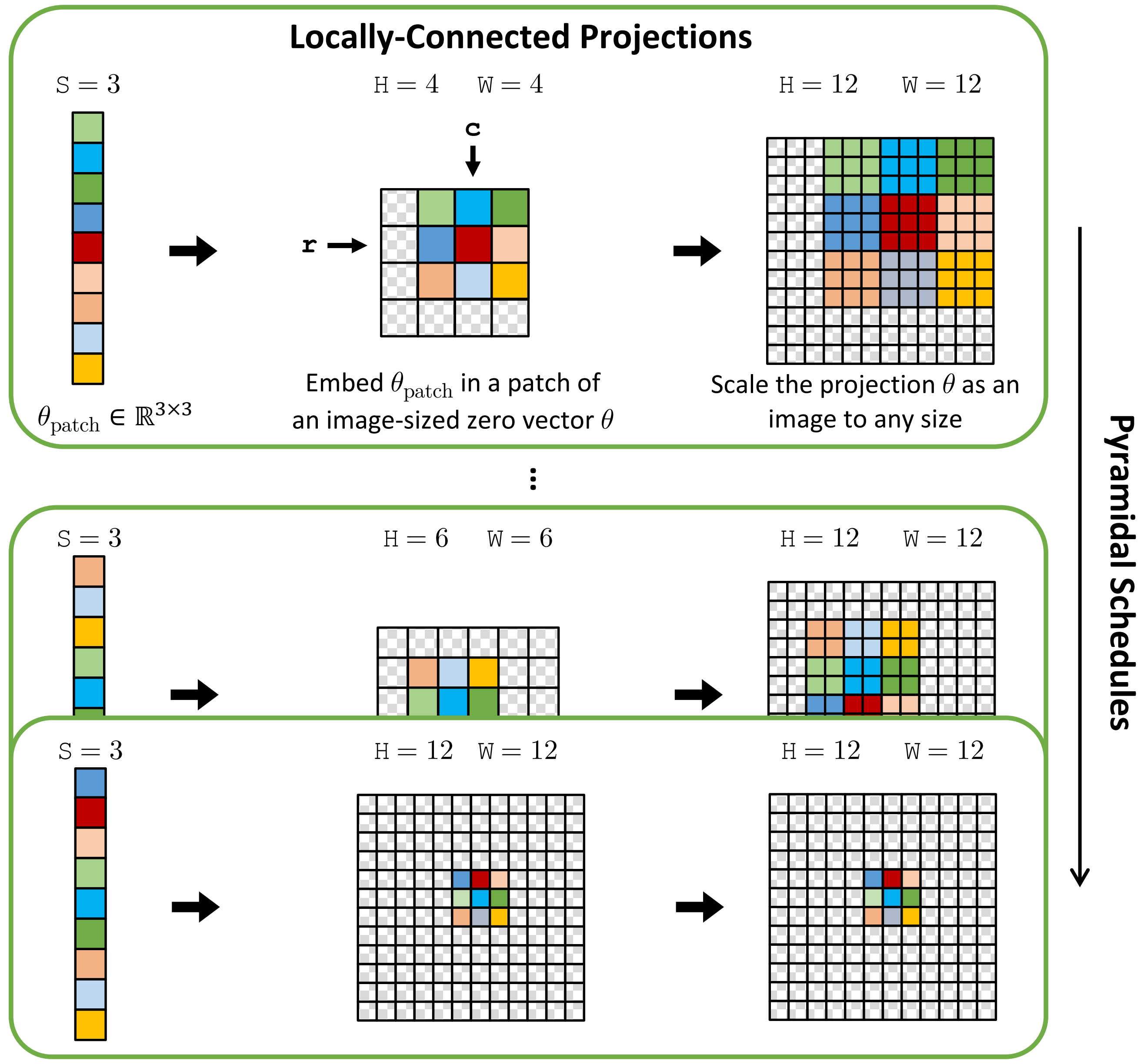
Intuition (Check the CDFs):



SWFs + Visual Inductive Biases

Motivated by the analogy of uniform projections and fully-connected layers

Change the distribution of the projections



Experiments

Unconditional Image Generation

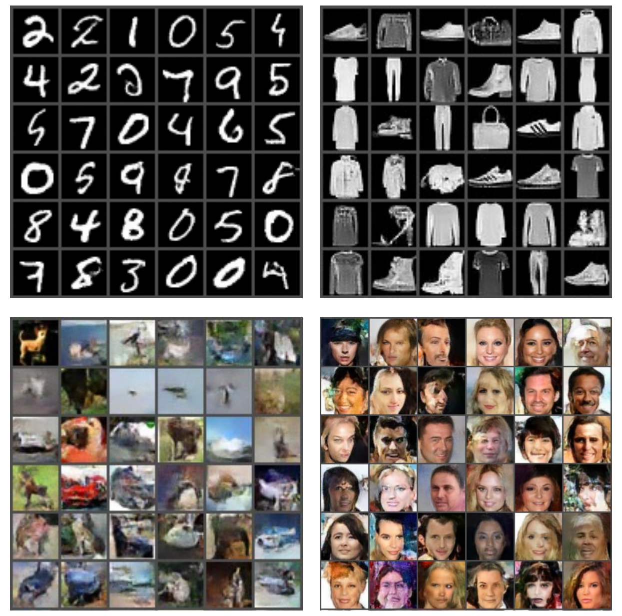


Table 1. FID↓ scores obtained by ℓ-SWF on CIFAR-10 and CelebA. ⋄ Use 160 × 160 center-cropping. ∗ Use 128 × 128 center-cropping. † Use 140 × 140 center-cropping.

Method	CIFAR-10	CelebA
Auto-encoder based		
VAE (Kingma & Welling, 2013)	155.7	85.7⋄
SWAE (Wu et al., 2019)	107.9	48.9*
WAE (Tolstikhin et al., 2017)	–	42†
CWAE (Knop et al., 2020)	120.0	49.7†
Autoregressive & Energy based		
PixelCNN (Van den Oord et al., 2016)	65.9	–
EBM (Du & Mordatch, 2019)	37.9	–
Adversarial		
WGAN (Arjovsky et al., 2017)	55.2	41.3⋄
WGAN-GP (Gulrajani et al., 2017)	55.8	30.0⋄
CSW (Nguyen & Ho, 2022b)	36.8	–
SWGAN (Wu et al., 2019)	17.0	13.2*
Score based		
NCSN (Song & Ermon, 2019)	25.3	–
Nonparametric		
SWF (Liutkus et al., 2019)	> 200	> 150†
SINF (Dai & Seljak, 2021)	66.5	37.3*
ℓ-SWF (Ours)	59.7	38.3†

Class-Conditional Image Generation

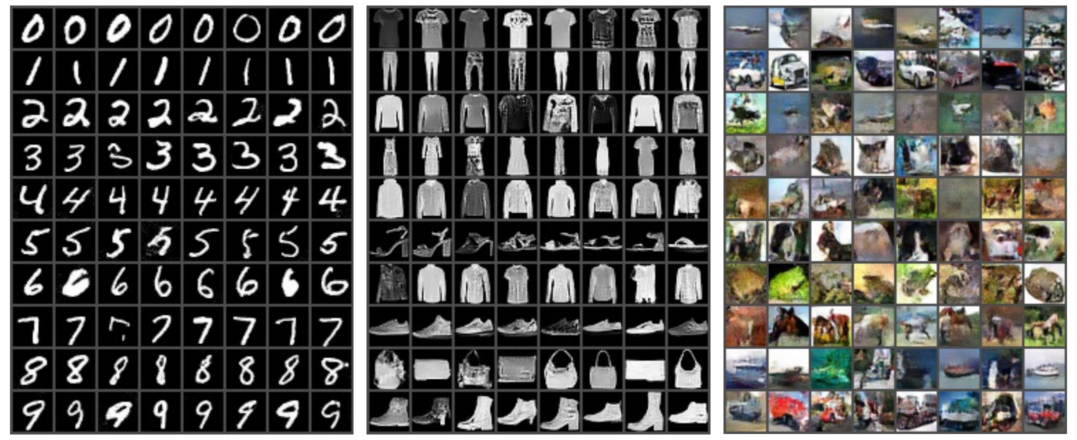
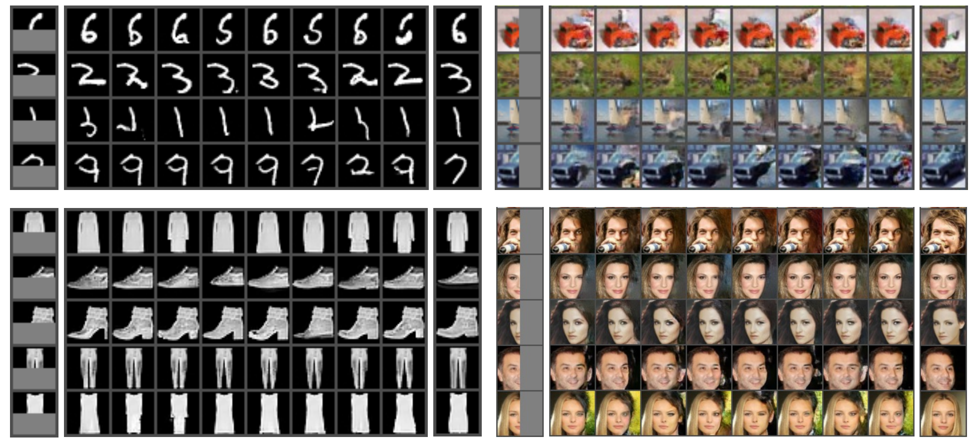
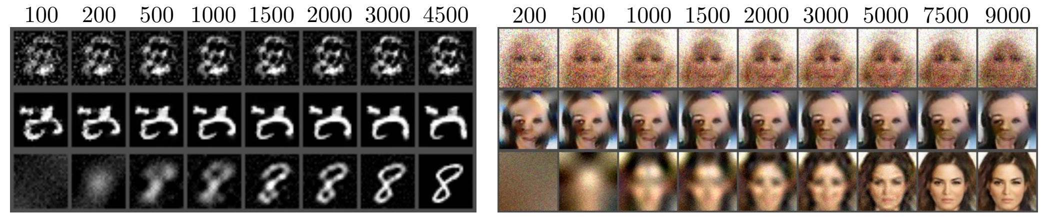


Image Inpainting



Locally-Connected Projections & Pyramidal Schedules

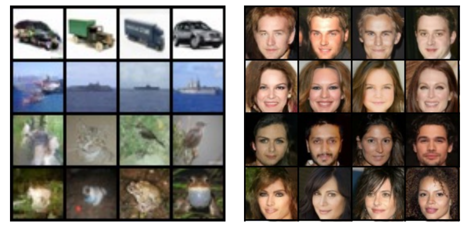


SWF (Liutkus et al., 2019)

+ Locally-Connected Projections

+ Pyramidal Schedules

Nearest Neighbors



Primary Contact: Chao Du (duchao0726@gmail.com); Code: https://github.com/duchao0726/Conditional-SWF