A. Details of Section 3.1: Benamou-Brenier formulation in Lagrangian coordinates

The Benamou-Brenier formulation of the optimal transportation (OT) problem in Eulerian coordinates is

$$\min_{\mathbf{f},\rho} \quad \int_0^T \int \|\mathbf{f}(\mathbf{x},t)\|^2 \rho_t(\mathbf{x})\,\mathrm{d}\mathbf{x}\mathrm{d}t \tag{18a}$$

$$\text{subject to} \quad \frac{\partial \rho_t}{\partial t} = -\operatorname{div}\left(\rho_t \mathbf{f}\right), \tag{18b}$$

$$\rho_0(\mathbf{x}) = p, \tag{18c}$$

$$\rho_T(\mathbf{z}) = q. \tag{18d}$$

The connection between continuous normalizing flows (CNF) and OT becomes transparent once we rewrite (18) in Lagrangian coordinates. Indeed, for regular enough velocity fields $\mathbf{f}$ one has that the solution of the continuity equation (18b), (18c) is given by $\rho_t = \mathbf{z}(\cdot,t)\sharp p$ where $\mathbf{z}$ is the flow

$$\dot{\mathbf{z}}(\mathbf{x},t) = \mathbf{f}(\mathbf{z}(\mathbf{x},t),t), \quad \mathbf{z}(\mathbf{x},0) = \mathbf{x}.$$

The relation $\rho_t = \mathbf{z}(\cdot,t)\sharp p$ means that for arbitrary test function ϕ we have that

$$\int \phi(\mathbf{x})\rho_t(\mathbf{x},t)d\mathbf{x} = \int \phi(\mathbf{z}(\mathbf{x},t))p(\mathbf{x})d\mathbf{x}$$

Therefore (18) can be rewritten as

$$\min_{\mathbf{f}} \quad \int_0^T \int \|\mathbf{f}(\mathbf{z}(\mathbf{x},t),t)\|^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x}\mathrm{d}t \tag{19a}$$

$$\text{subject to} \quad \dot{\mathbf{z}}(\mathbf{x},t) = \mathbf{f}(\mathbf{z}(\mathbf{x},t),t), \tag{19b}$$

$$\mathbf{z}(\mathbf{x},0) = \mathbf{x}, \tag{19c}$$

$$\mathbf{z}(\cdot,T)\sharp p = q. \tag{19d}$$

Note that ρ_t is eliminated in this formulation. The terminal condition (18d) is trivial to implement in Eulerian coordinates (grid-based methods) but not so simple in Lagrangian ones (19d) (grid-free methods). To enforce (19d) we introduce a penalty term in the objective function that measures the deviation of $\mathbf{z}(\cdot,T)\sharp p$ from q. Thus, the penalized objective function is

$$\int_0^T \int \|\mathbf{f}(\mathbf{z}(\mathbf{x},t),t)\|^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x}\mathrm{d}t + \frac{1}{\lambda}\,\mathrm{KL}(\mathbf{z}(\cdot,T)\sharp p \,||\, q), \tag{20}$$

where $\lambda > 0$ is the penalization strength. Next, we observe that this objective function can be written as an expectation with respect to $\mathbf{x} \sim p$. Indeed, the Kullback-Leibler divergence is invariant under coordinate transformations, and therefore

$$\begin{aligned}\mathrm{KL}(\mathbf{z}(\cdot,T)\sharp p \,||\, q) &= \mathrm{KL}(p \,||\, \mathbf{z}^{-1}(\cdot,T)\sharp q) = \mathrm{KL}(p \,||\, p_\theta) \\ &= \mathop{\mathbb{E}}_{\mathbf{x}\sim p} \log \frac{p(\mathbf{x})}{p_\theta(\mathbf{x})} \\ &= \mathop{\mathbb{E}}_{\mathbf{x}\sim p} \log p(\mathbf{x}) - \mathop{\mathbb{E}}_{\mathbf{x}\sim p} \log p_\theta(\mathbf{x})\end{aligned}$$

Hence, multiplying the objective function in (20) by λ and ignoring the $\mathbf{f}$-independent term $\mathbb{E}_{\mathbf{x}\sim p} \log p(\mathbf{x})$ we obtain an equivalent objective function

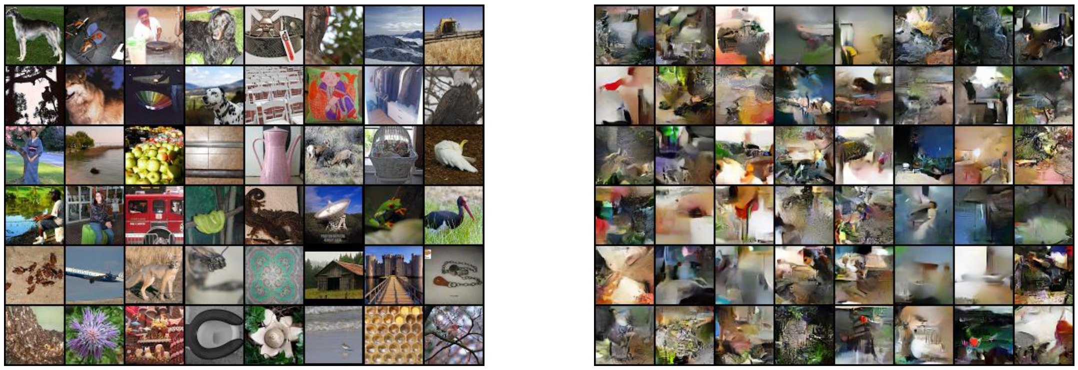
$$\mathop{\mathbb{E}}_{\mathbf{x}\sim p}\left\{\lambda \int_0^T \|\mathbf{f}(\mathbf{z}(\mathbf{x},t),t)\|^2\,\mathrm{d}t - \log p_\theta(\mathbf{x})\right\} \tag{21}$$

Finally, if we assume that $\{\mathbf{x}_i\}_{i=1}^N$ are iid sampled from p, we obtain the empirical objective function

$$\frac{\lambda}{N}\sum_{i=1}^N \int_0^T \|\mathbf{f}(\mathbf{z}(\mathbf{x}_i,t),t)\|^2\,\mathrm{d}t - \frac{1}{N}\sum_{i=1}^N \log p_\theta(\mathbf{x}_i) \tag{22}$$

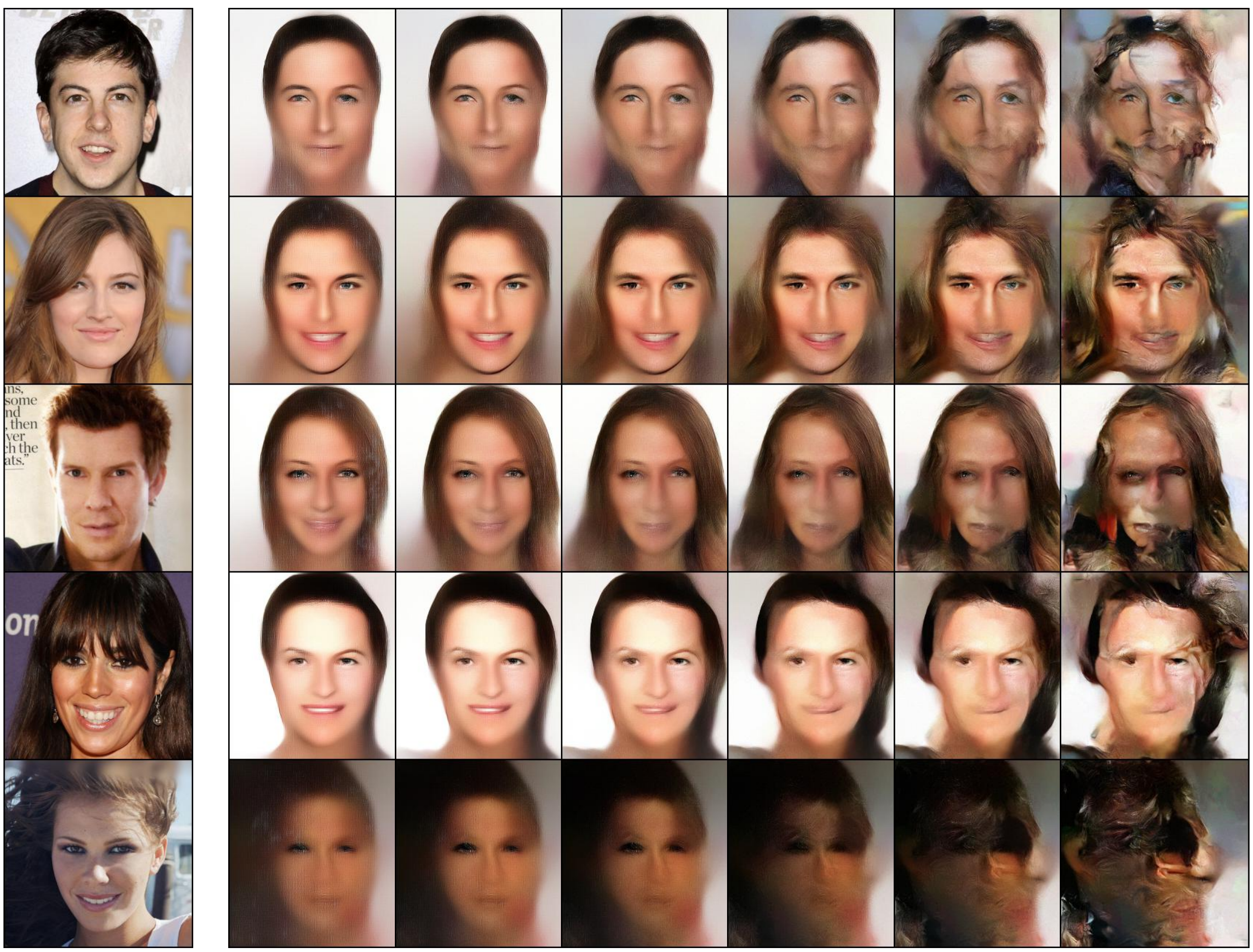
B. Additional results

Here we present additional generated samples on the two larger datasets considered, CelebA-HQ and ImageNet64. In addition bits/dim on clean images are reported in Table 2.



(a) real (b) generated

Figure 7. Quality of FFJORD RNODE generated images on ImageNet-64.



(a) real (b) $T = 0.5$ (c) $T = 0.6$ (d) $T = 0.7$ (e) $T = 0.8$ (f) $T = 0.9$ (g) $T = 1$

Figure 8. Quality of FFJORD RNODE generated images on CelebA-HQ. We use temperature annealing, as described in (Kingma & Dhariwal, 2018), to generate visually appealing images, with $T = 0.5, \ldots, 1$.

Table 2. Additional results and model statistics of FFJORD RNODE. Here we report validation bits/dim on both validation images, and on validation images with uniform variational dequantization (ie perturbed by uniform noise). We also report number of trainable model parameters.

DATASET	BITS/DIM (CLEAN)	BITS/DIM (DIRTY)	# PARAMETERS
MNIST	0.92	0.97	8.00e5
CIFAR10	3.25	3.38	1.36e6
IMAGENET64	3.72	3.83	2.00e6
CELEBA-HQ256	0.72	1.04	4.61e6