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

$$\min_{\mathbf{f}, \rho} \int_0^T \int ||\mathbf{f}(\mathbf{x}, t)||^2 \rho_t(\mathbf{x}) \, d\mathbf{x} \, dt$$  \hspace{1cm} (18a)

subject to

$$\frac{\partial \rho_t}{\partial t} = - \text{div} (\rho_t \mathbf{f}),$$ \hspace{1cm} (18b)

$$\rho_0(\mathbf{x}) = p,$$ \hspace{1cm} (18c)

$$\rho_T(\mathbf{z}) = q.$$ \hspace{1cm} (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 $\rho_t = z(\cdot, t) \rho$ where $z$ is the flow

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

The relation $\rho_t = z(\cdot, t) \rho$ means that for arbitrary test function $\phi$ we have that

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

Therefore (18) can be rewritten as

$$\min_{\mathbf{f}} \int_0^T \int ||\mathbf{f}(z(\mathbf{x}, t), t)||^2 \rho(\mathbf{x}) \, d\mathbf{x} \, dt$$  \hspace{1cm} (19a)

subject to

$$\dot{z}(\mathbf{x}, t) = f(z(\mathbf{x}, t), t),$$ \hspace{1cm} (19b)

$$z(\mathbf{x}, 0) = \mathbf{x},$$ \hspace{1cm} (19c)

$$z(\cdot, T) \rho = q.$$ \hspace{1cm} (19d)

Note that $\rho_t$ is eliminated in this formulation. The terminal condition (19d) 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 $z(\cdot, T) \rho$ from $q$. Thus, the penalized objective function is

$$\int_0^T \int ||\mathbf{f}(z(\mathbf{x}, t), t)||^2 \rho(\mathbf{x}) \, d\mathbf{x} \, dt + \frac{1}{\lambda} \text{KL}(z(\cdot, T) \rho \| q),$$  \hspace{1cm} (20)

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

$$\text{KL}(z(\cdot, T) \rho \| q) = \text{KL}(p \| z^{-1}(\cdot, T) \rho q) = \text{KL}(p \| \rho)$$

$$= \mathbb{E}_{\mathbf{x} \sim \rho} \log \frac{p(\mathbf{x})}{\rho(\mathbf{x})}$$

$$= \mathbb{E}_{\mathbf{x} \sim \rho} \log p(\mathbf{x}) - \mathbb{E}_{\mathbf{x} \sim \rho} \log \rho(\mathbf{x})$$

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

$$\mathbb{E}_{\mathbf{x} \sim \rho} \left\{ \lambda \int_0^T ||\mathbf{f}(z(\mathbf{x}, t), t)||^2 \, dt - \log p_0(\mathbf{x}) \right\}$$  \hspace{1cm} (21)

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

$$\frac{\lambda}{N} \sum_{i=1}^N \int_0^T ||\mathbf{f}(z(x_i, t), t)||^2 \, dt - \frac{1}{N} \sum_{i=1}^N \log p_0(x_i)$$  \hspace{1cm} (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.
How to Train Your Neural ODE: the World of Jacobian and Kinetic Regularization

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

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.

<table>
<thead>
<tr>
<th>DATASET</th>
<th>BITS/DIM (CLEAN)</th>
<th>BITS/DIM (DIRTY)</th>
<th># PARAMETERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>0.92</td>
<td>0.97</td>
<td>8.00e5</td>
</tr>
<tr>
<td>CIFAR10</td>
<td>3.25</td>
<td>3.38</td>
<td>1.36e6</td>
</tr>
<tr>
<td>IMAGENET64</td>
<td>3.72</td>
<td>3.83</td>
<td>2.00e6</td>
</tr>
<tr>
<td>CELEBA-HQ256</td>
<td>0.72</td>
<td>1.04</td>
<td>4.61e6</td>
</tr>
</tbody>
</table>