TO-FLOW: Efficient Continuous Normalizing Flows with Temporal Optimization adjoint with Moving Speed

Supplementary Material

A. Generic form of Temporal Optimization

If optimizing t_0 and T at the same time, then the objective function has the following form:

$$\begin{aligned}
&\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta},t_0\in\mathbb{R},T\in\mathbb{R}} L(\boldsymbol{\theta},t_0,T)\\
&= -\mathbb{E}_{p_{\mathbf{x}}}\{\log p(\mathbf{z}(t_0);\boldsymbol{\theta})\}\\
&= -\mathbb{E}_{p_{\mathbf{x}}}\left\{\log p(\mathbf{z}(T);\boldsymbol{\theta}) + \int_{t_0}^{T} Tr(\mathbf{J}(t,\boldsymbol{\theta}))dt\right\},
\end{aligned} \tag{1}$$

The optimization of T has been given above, hence in this section we only need to show the derivation of $\frac{\partial L}{\partial t_0}$. Also, it's not easy to get the $\frac{\partial L}{\partial t_0}$ and $\log p(\mathbf{z}(t_0))$ are unknown, so we rewrite the formula and introduce the chain rule by using intermediate variable $\mathbf{z}(t_0)$ for simplicity. Then we get

$$\begin{aligned}
&\frac{\partial L(\boldsymbol{\theta},t_0,T)}{\partial t_0}\\
&= -\frac{\partial \mathbb{E}_{p_{\mathbf{x}}}\{\log p(\mathbf{z}(t_0),\boldsymbol{\theta})\}}{\partial t_0}\\
&= -\frac{\partial \mathbb{E}_{p_{\mathbf{x}}}\{\log p(\mathbf{z}(T),\boldsymbol{\theta})\}}{\partial t_0} + \mathbb{E}_{p_{\mathbf{x}}}\{Tr(\mathbf{J}(t_0,\boldsymbol{\theta}))\}\\
&= -\mathbb{E}_{p_{\mathbf{x}}}\left\{\frac{\partial \log p(\mathbf{z}(T),\boldsymbol{\theta})}{\partial \mathbf{z}(t_0)} \circ \frac{\partial \mathbf{z}(t_0)}{\partial t_0}\right\} + \mathbb{E}_{p_{\mathbf{x}}}\{Tr(\mathbf{J}(t_0,\boldsymbol{\theta}))\}.
\end{aligned} \tag{2}$$

Once we get $\frac{\partial L}{\partial t_0}$, we put $(\frac{\partial L}{\partial t_0}, t_0)$ into the temporal optimizer Q as mentioned in the text and optimize (t_0, T) together.

B. Comparison of 2-dimensional images at different stages

We compare the generated images of FFJORD and TO-FLOW for eight 2D datasets at 1000, 3000, 5000, 7000 and 9000 iterations. The results are shown in Figures 1, 2, 3, 4, 5, 6 and 7.

C. Images under different temporal regularization

We generate images of the three datasets under different temporal regularization. The results are shown in Figures 8, 9, 10, 11, 12, 13.

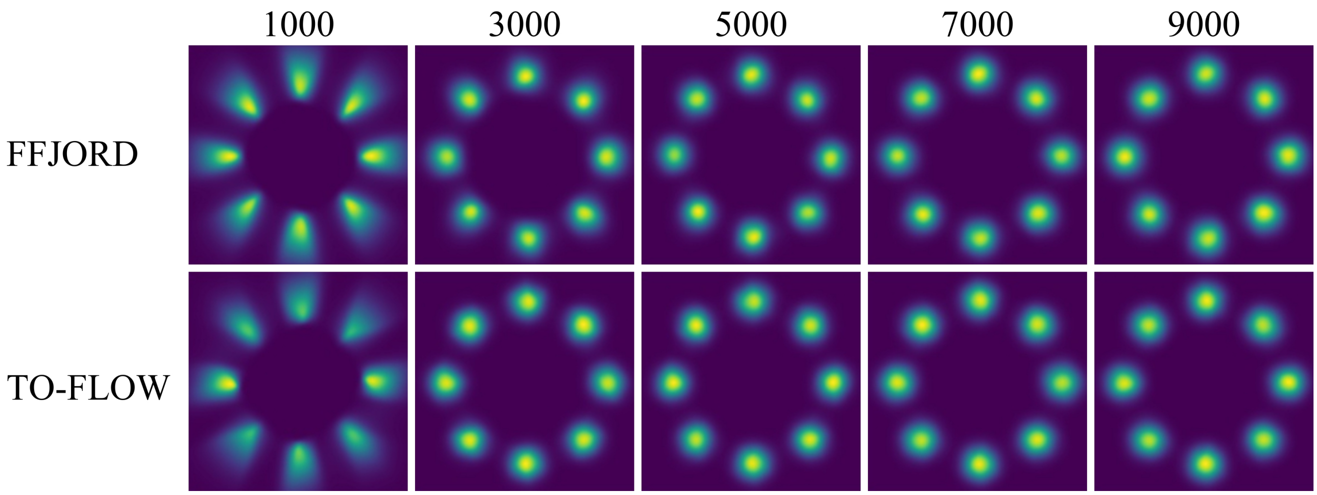

Figure 1. Comparison of FFJORD and TO-FLOW on 8gaussian data set. The numbers at the top of the images represent the number of iterations of the model.

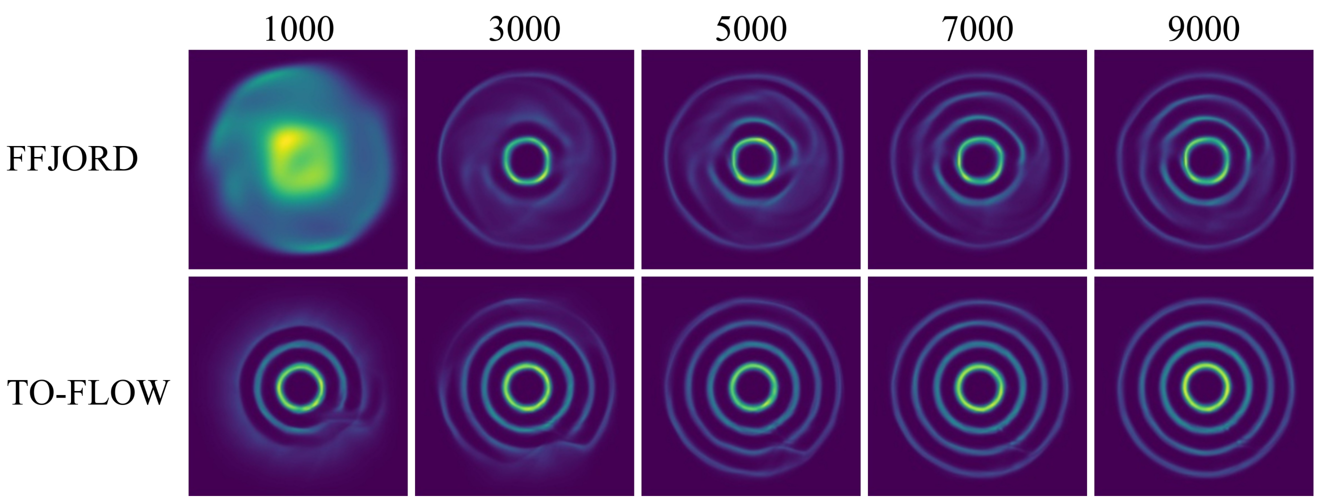

Figure 2. Comparison of FFJORD and TO-FLOW on rings data set. The numbers at the top of the images represent the number of iterations of the model.

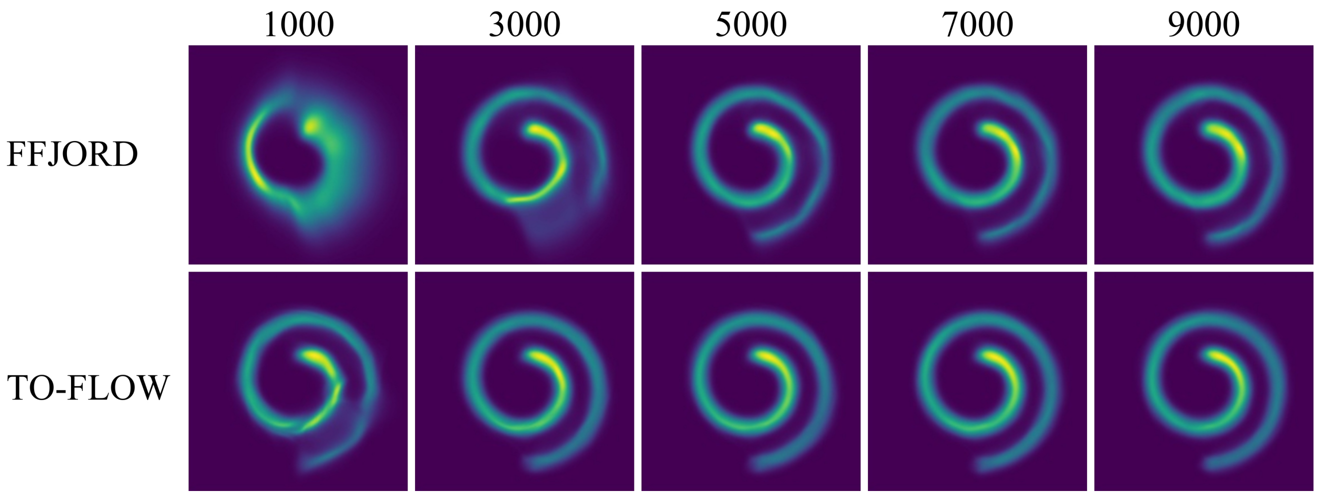

Figure 3. Comparison of FFJORD and TO-FLOW on swissroll data set. The numbers at the top of the images represent the number of iterations of the model.

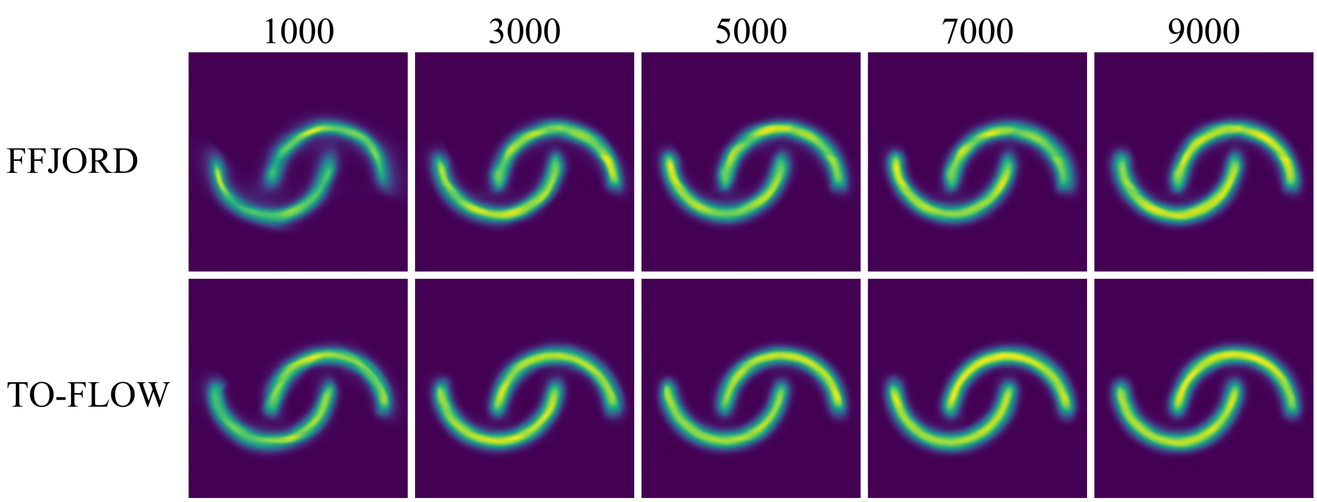

Figure 4. Comparison of FFJORD and TO-FLOW on moons data set. The numbers at the top of the images represent the number of iterations of the model.

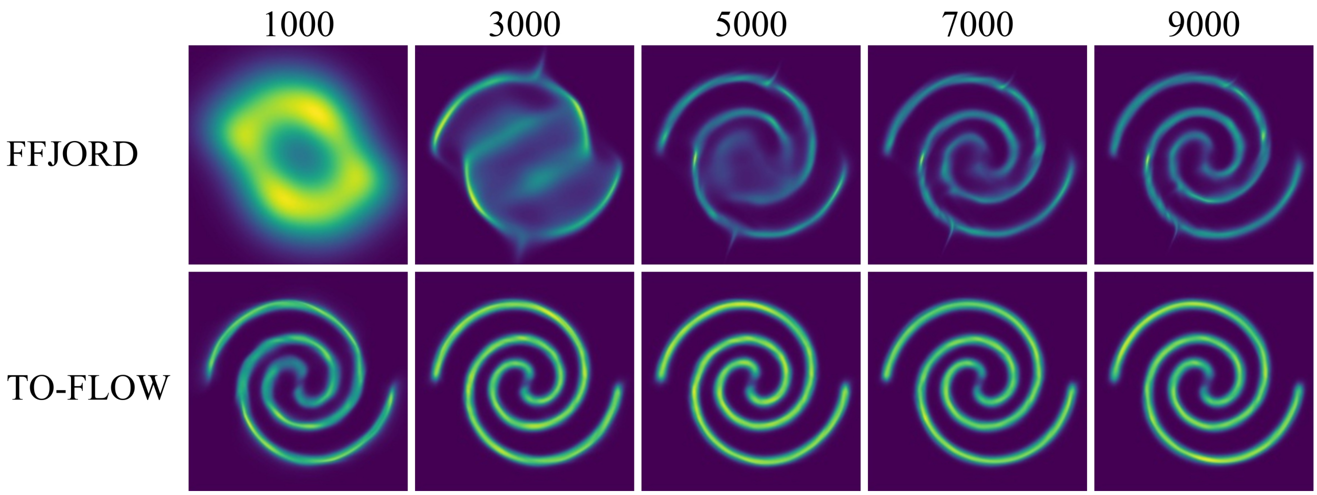

Figure 5. Comparison of FFJORD and TO-FLOW on 2spirals data set. The numbers at the top of the images represent the number of iterations of the model.

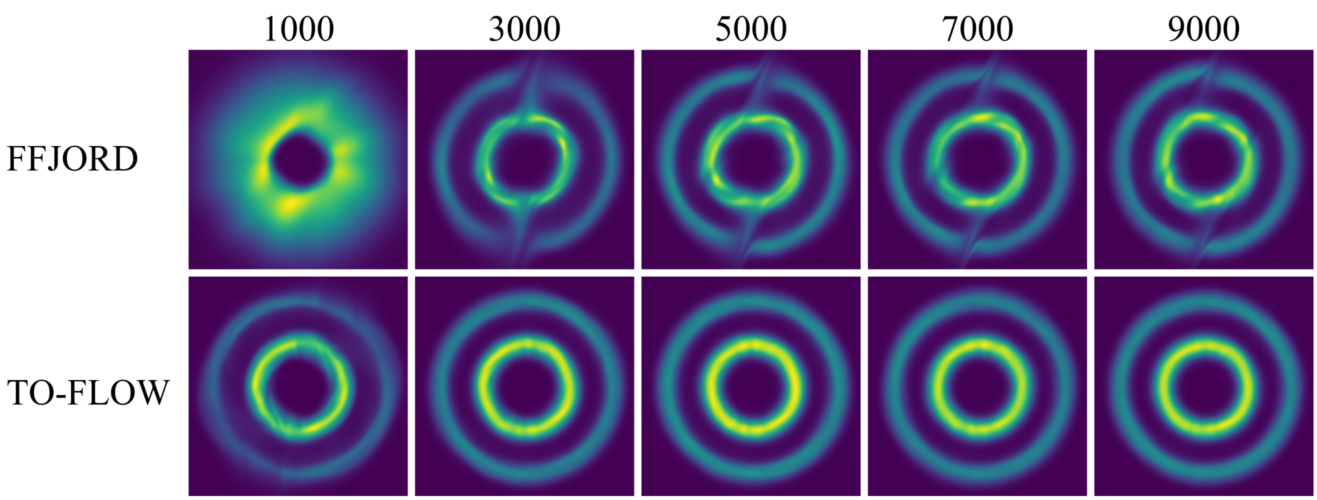

Figure 6. Comparison of FFJORD and TO-FLOW on circles data set. The numbers at the top of the images represent the number of iterations of the model.

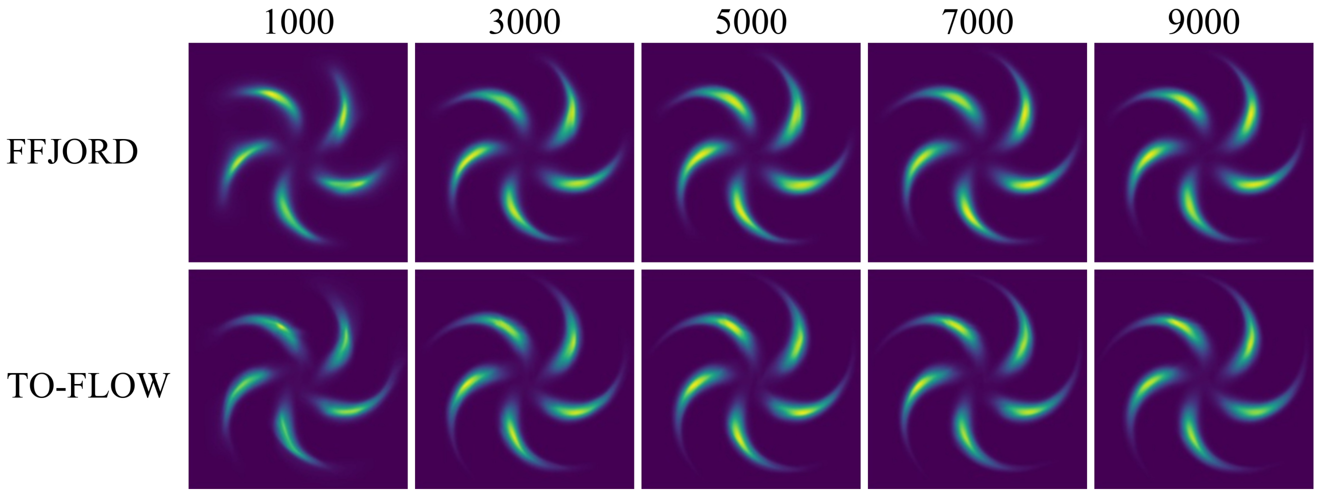

Figure 7. Comparison of FFJORD and TO-FLOW on pinwheel data set. The numbers at the top of the images represent the number of iterations of the model.

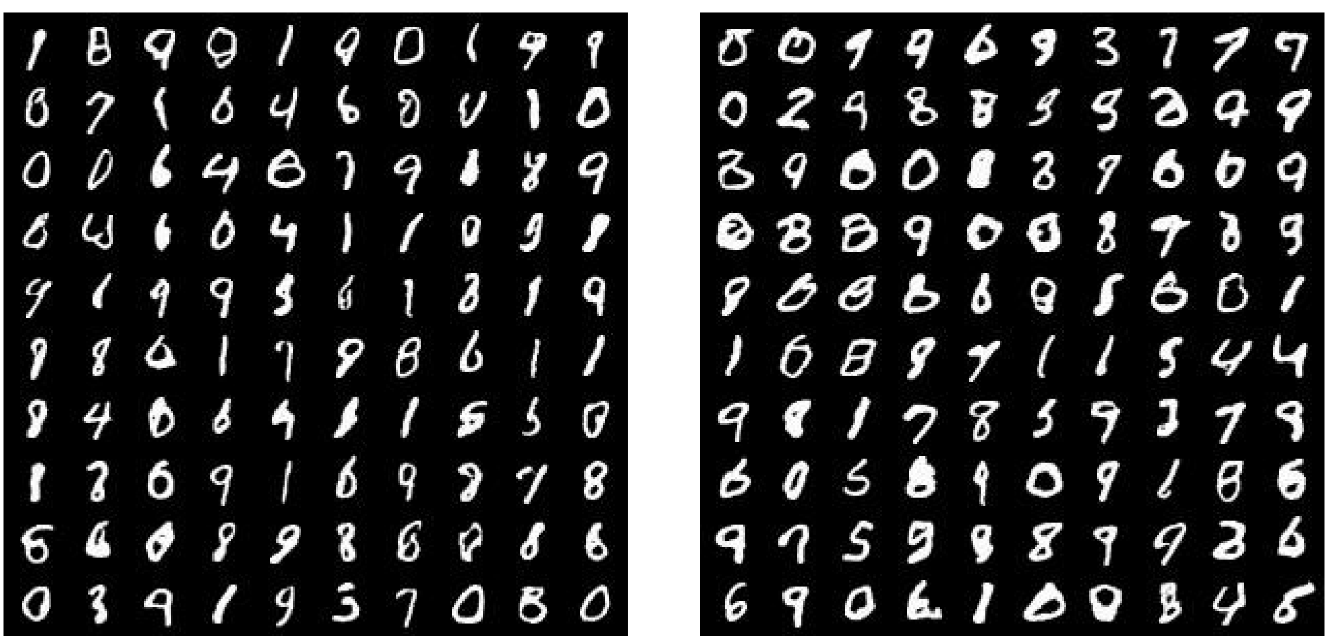

Figure 8. Samples from MNIST.
Left: $\alpha = 0$ Right: $\alpha = 0.1$

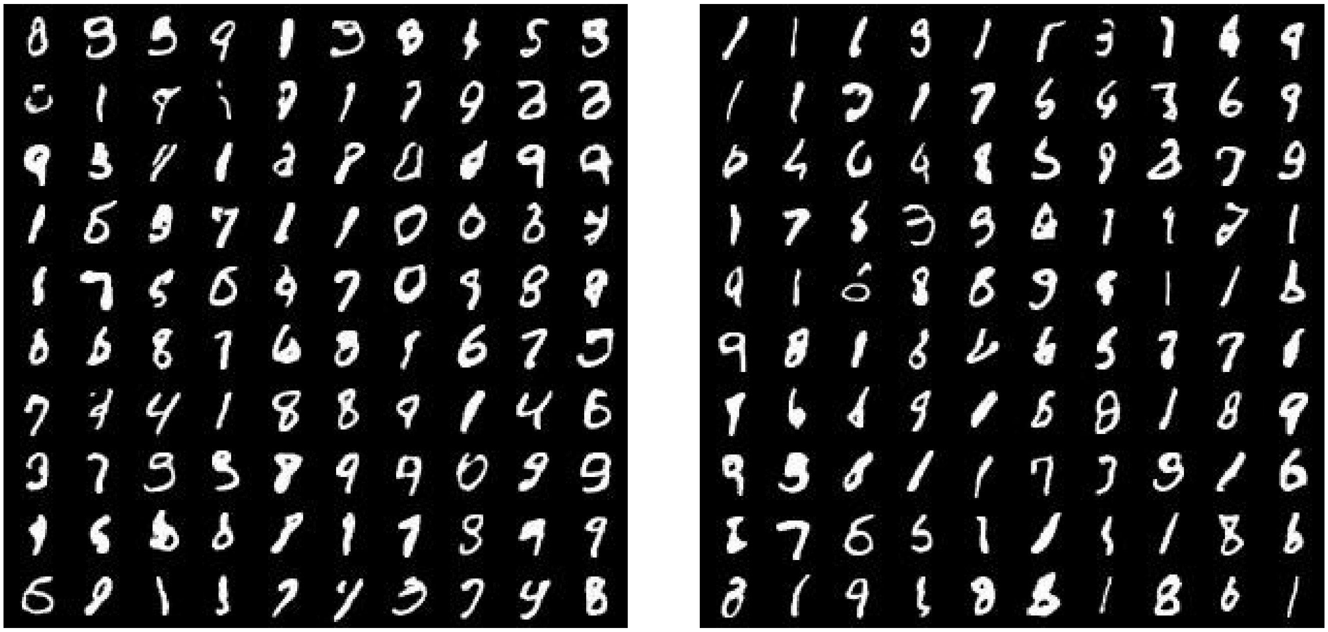

Figure 9. Samples from MNIST.
Left: $\alpha = 0.2$ Right: $\alpha = 0.3$

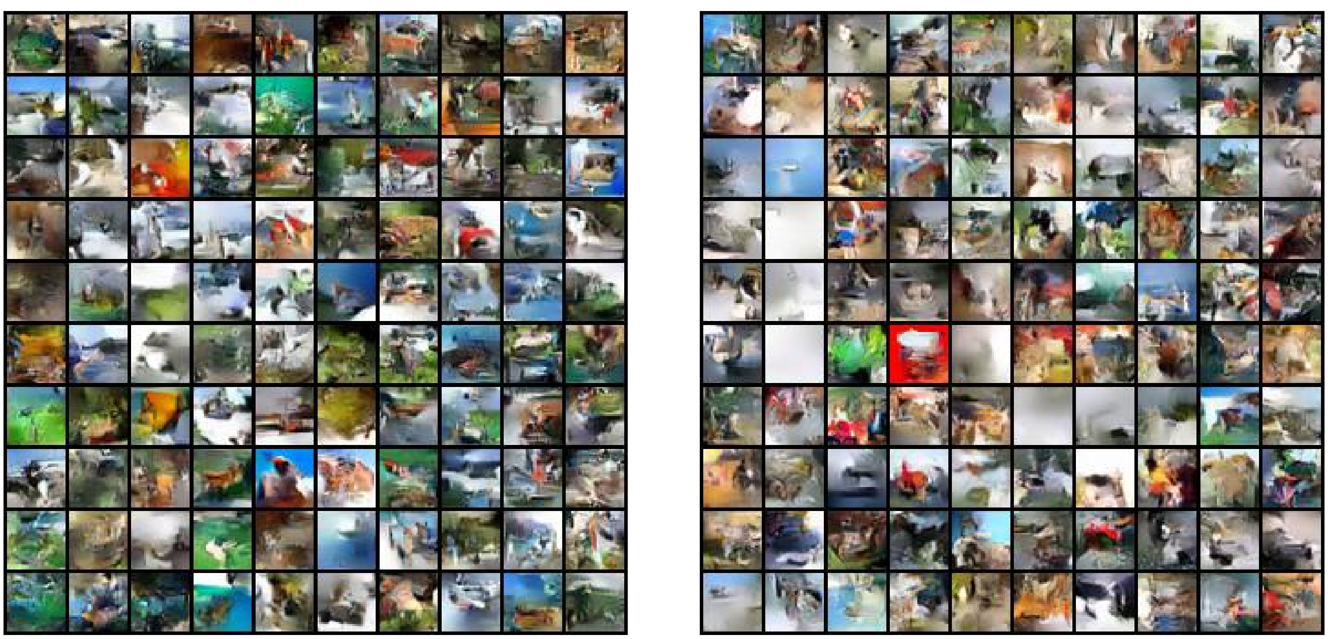

Figure 10. Samples from CIFAR-10.
Left: $\alpha = 0$ Right: $\alpha = 0.1$

Figure 11. Samples from CIFAR-10.
Left: $\alpha = 0.2$ Right: $\alpha = 0.3$

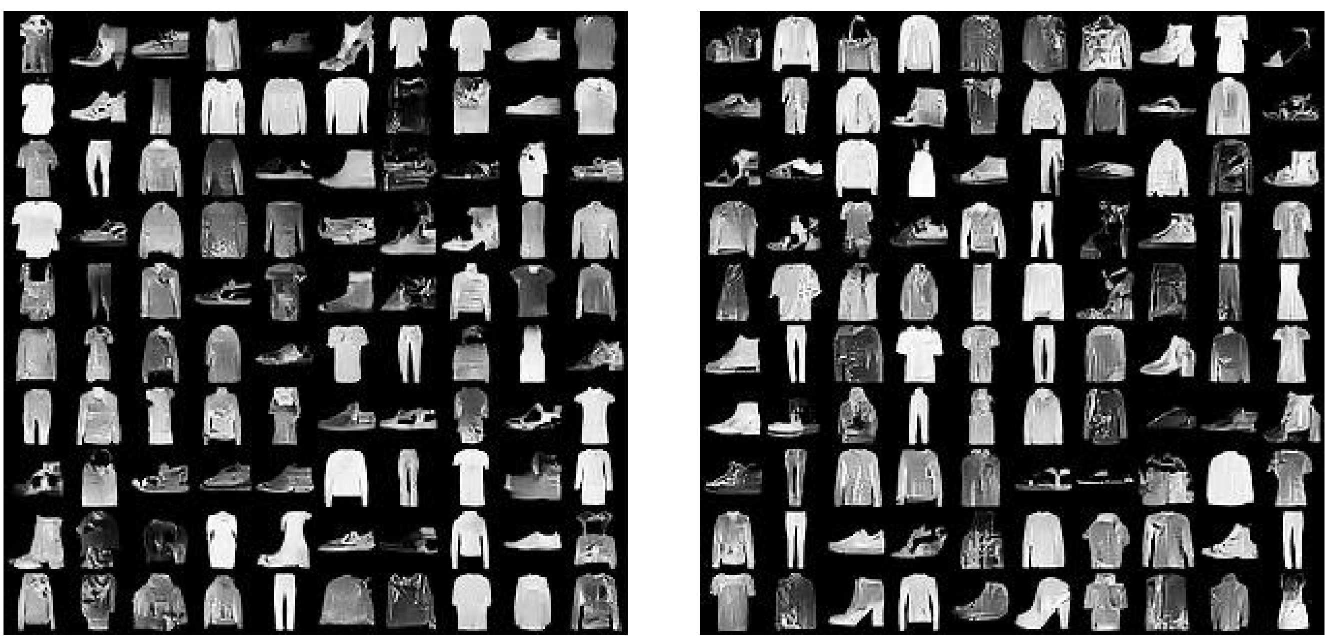

Figure 12. Samples from FASHION-MNIST.
Left: $\alpha = 0$ Right: $\alpha = 0.1$

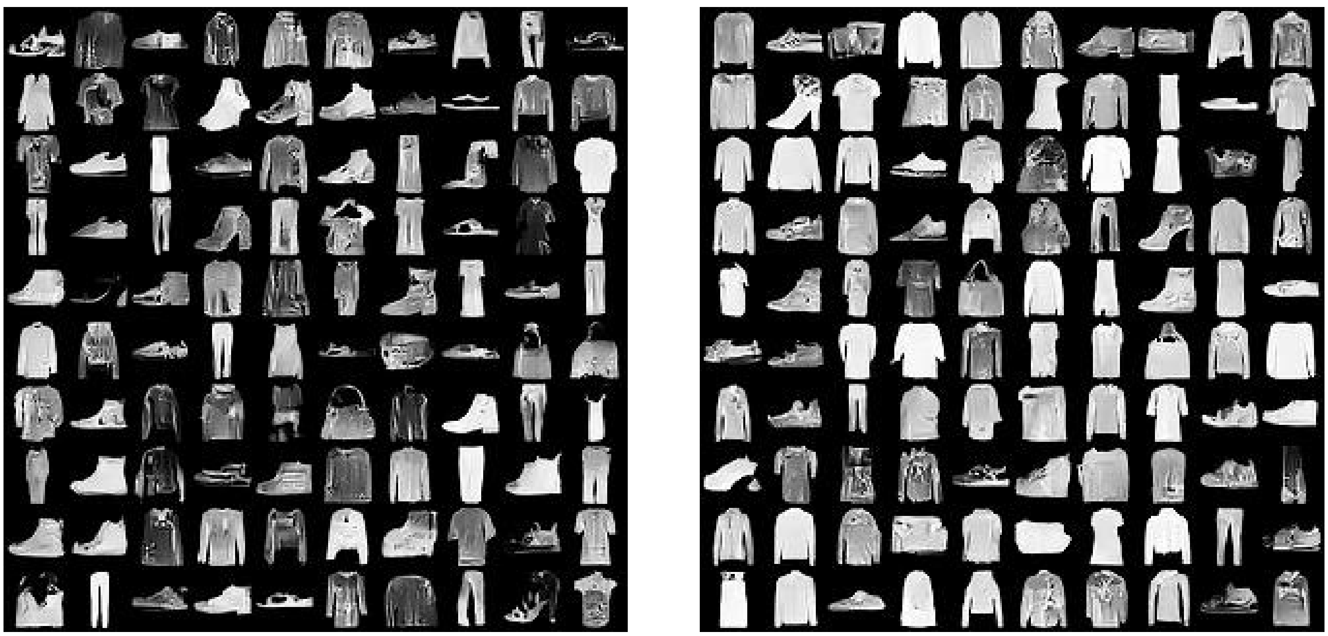

Figure 13. Samples from FASHION-MNIST.
Left: $\alpha = 0.2$ Right: $\alpha = 0.3$