RealNVP

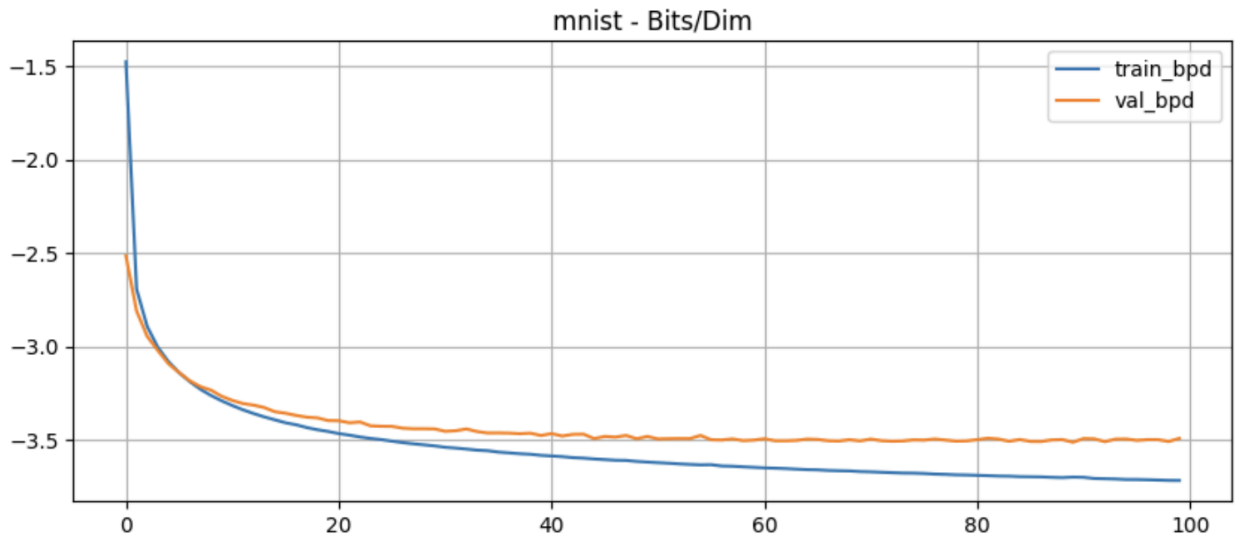
GPU - A100 GPU was used

The model converged stably on both datasets, with fast improvement in early epochs.
MNIST achieved better results due to simpler image structure.
FashionMNIST was harder to model, with noisier generated samples.

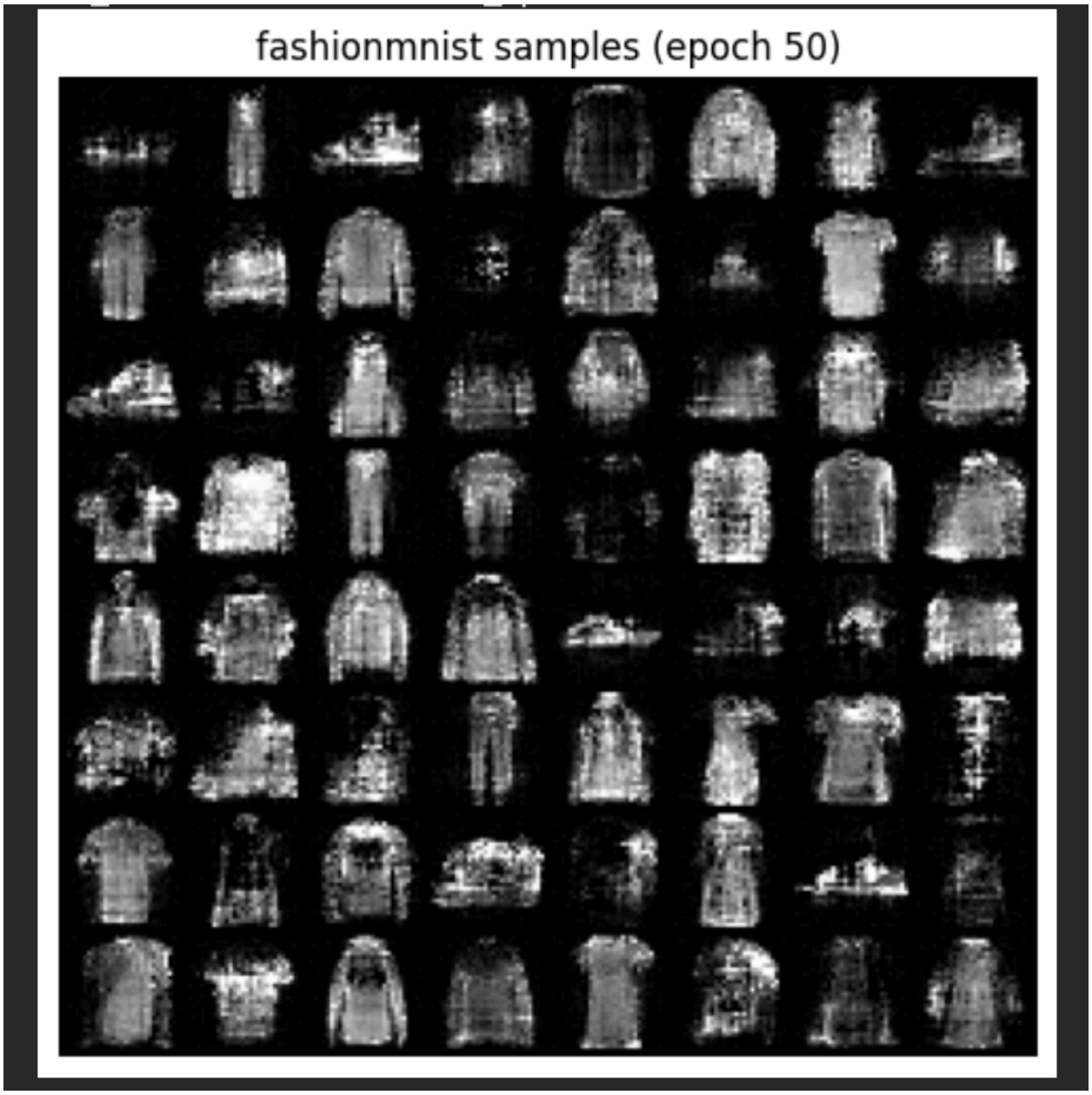
Generated MNIST samples were recognizable but slightly noisy; FashionMNIST samples were blurrier. Overal, both generated examples are far away from ideal

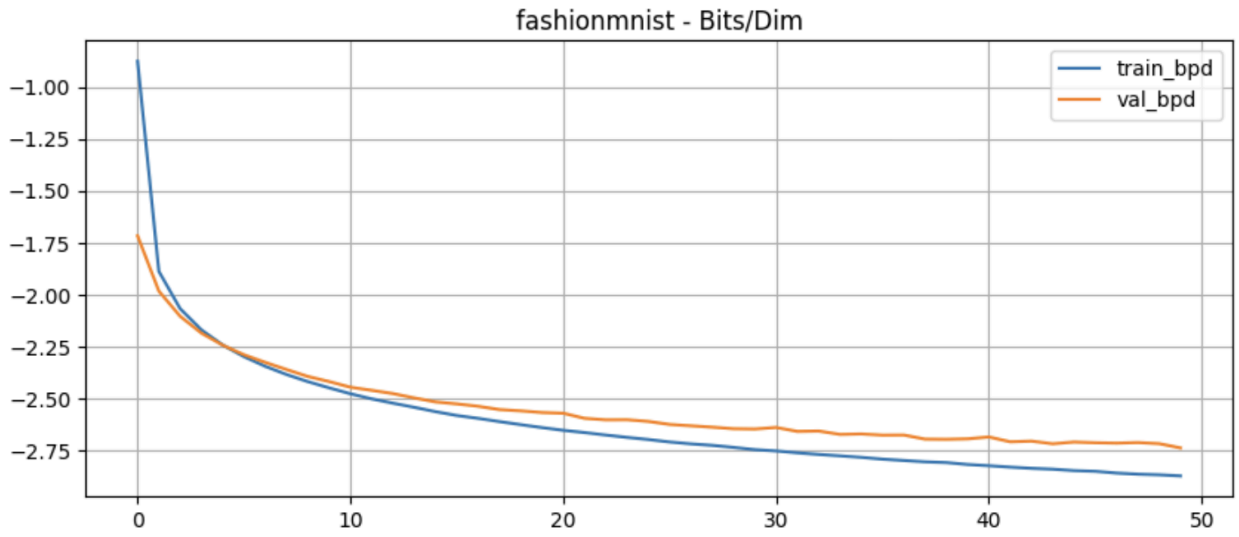
Likelihood improved faster than visual quality, which is typical for flow-based models.

MNIST example


mnist - Bits/Dim
train_bpd
val_bpd
−1.5
−2.0
−2.5
−3.0
−3.5
0
20
40
60
80
100

FashionMNIST example





The results might not be optimal due to limited model depth or a relatively small number of coupling layers.

I think the results can be improved by increasing the number of flow steps or using multi-scale RealNVP architectures. Training for more epochs I don't think will help, I experimented with epochs.