Model	NLL $\pi = 0$*	NLL $\pi = 0.5$	NLL $\pi = 1$
NCPN (ML)	2.25	3.72	4.35
NCPN (NCML VE)	2.22	3.63	4.21
NPCN (NCML sub-VP)	2.31	3.44	3.98
NCPN (NCML VP)	2.48	3.14	3.67

Table 2: Results on CelebA 64x64. Negative log-likelihood (NLL) is in bits per dimension. Lower is better. *NLL with $\pi = 0$ is equivalent to NLL of the original data.

A Appendix

A.1 Additional Experimental Details

We compare against Kingma et al. (2021); Song et al. (2021); Child (2020); Ho et al. (2019); Grcić et al. (2021); Van Den Oord et al. (2016); Chen et al. (2018); Child et al. (2019). Some results could not be included due to the irreproducibility of the techniques.

For our NCML-trained models, the diffusion times of the VE, VP, and sub-VP SDEs were chosen to be $t = 0.5$, $t = 0.1$, and $t = 0.025$, respectively. The values are somewhat arbitrary, but chosen such that the standard deviation of the per-pixel differences between samples in p_{data} and their noised counterparts in p_T was ≈ 10. We suspect that further improvements can be made to the empirical results if these numbers were chosen more judiciously.

A.2 Additional Figures and Tables

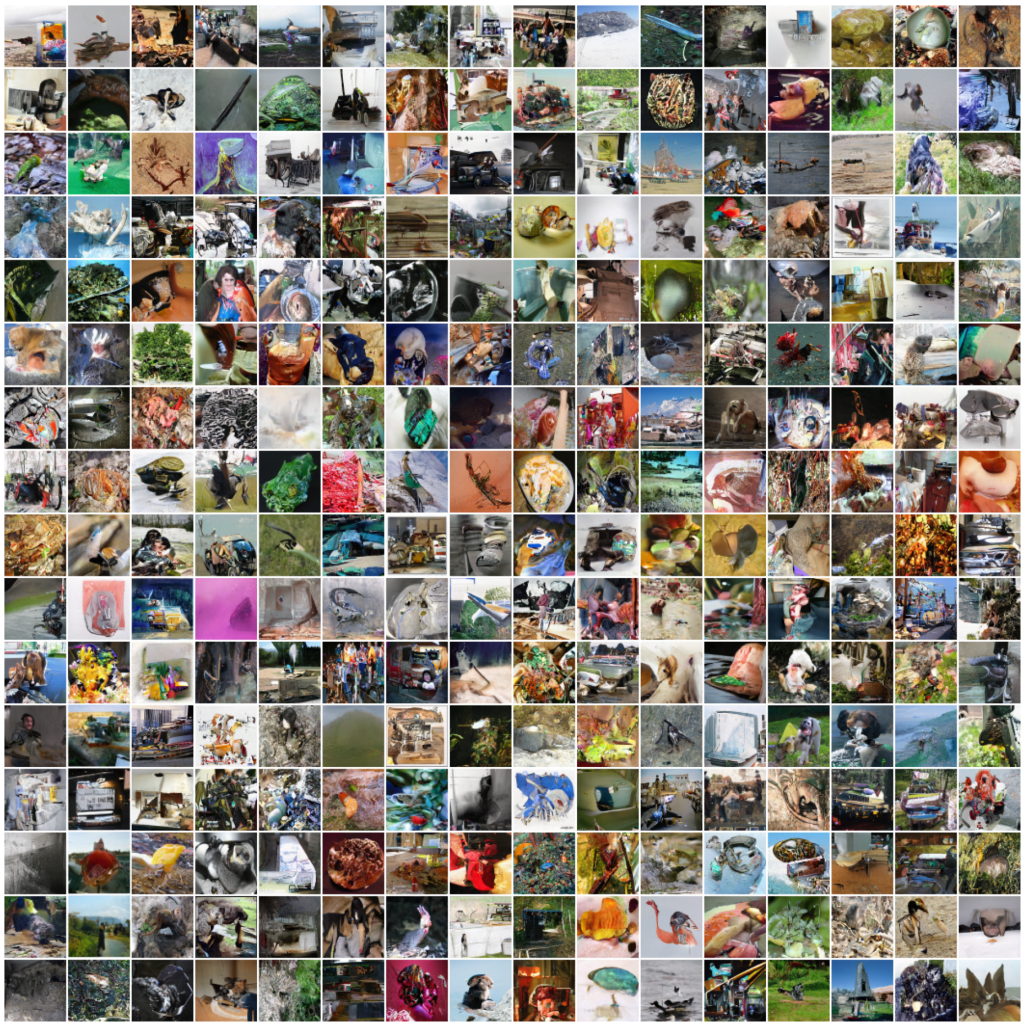

Figure 6: Samples from NCPN trained on ImageNet 64x64, with p_t as a variance preserving (VP) diffusion process.

Figure 7: Samples from NCPN trained on CelebA 64x64, with p_t as a variance preserving (VP) diffusion process.

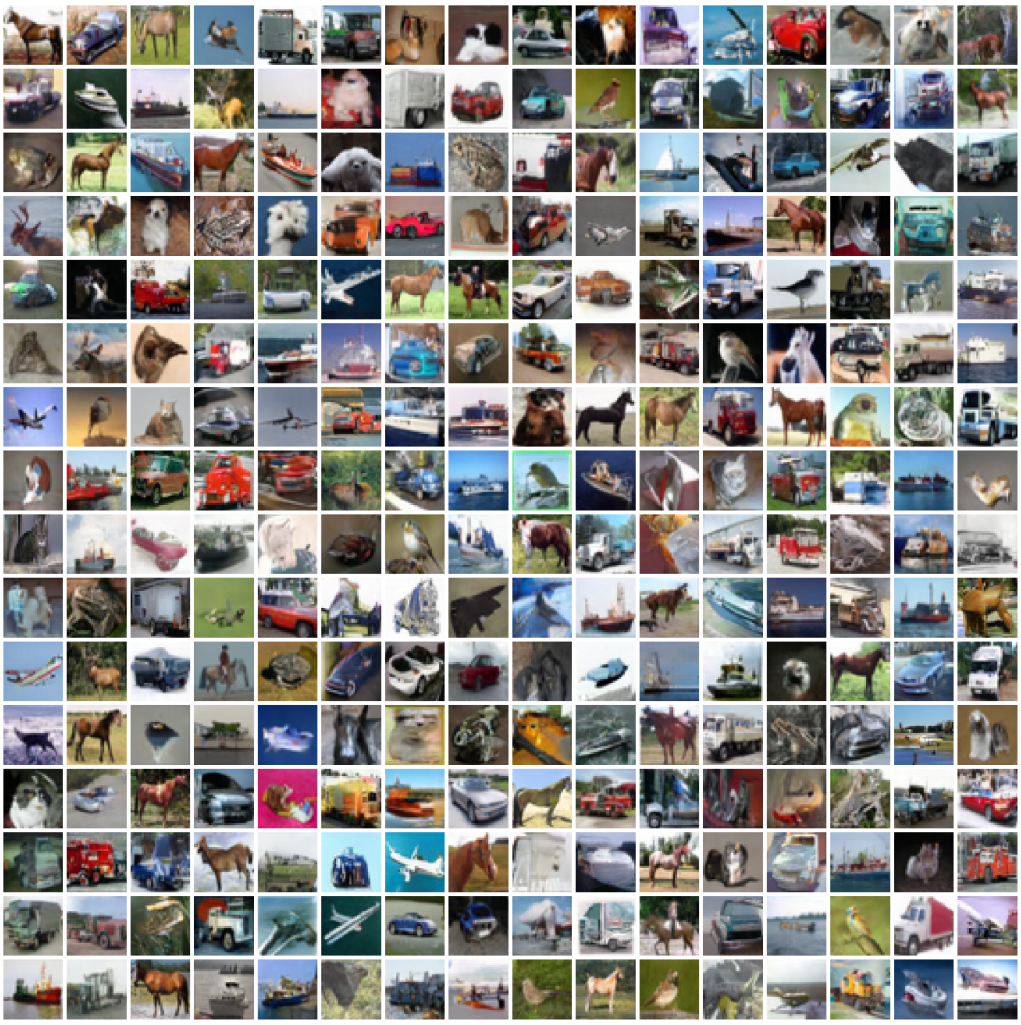

Figure 8: Samples from NCPN trained on CIFAR-10, with p_t as a variance preserving (VP) diffusion process.

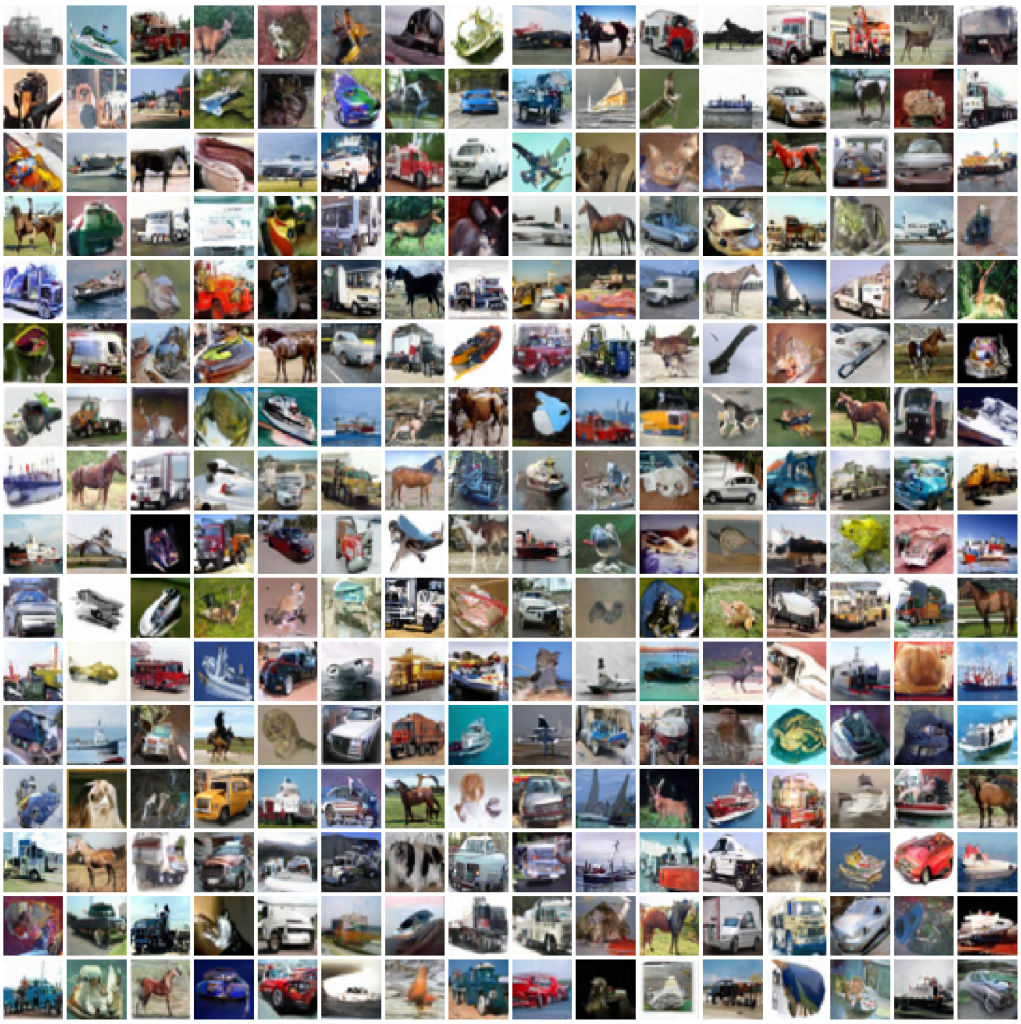

Figure 9: Samples from NCPN trained on CIFAR-10, with p_t as a sub-variance preserving (sub-VP) diffusion process.

Figure 10: Samples from NCPN trained on CIFAR-10, with p_t as a variance exploding (VE) diffusion process.