Appendix

A. Ablation

Dataset Pair + Model Specifics	AUROC scores
FashionMNIST/MNIST + L1 + $w(t)$	**1.000**
FashionMNIST/MNIST + L2 + $w(t)$	0.966
FashionMNIST/MNIST + L1	0.989
FashionMNIST/MNIST + L2	0.934
CIFAR10/SVHN + L1 + $w(t)$	**0.964**
CIFAR10/SVHN + L2 + $w(t)$	0.955
CIFAR10/SVHN + L1	0.963
CIFAR10/SVHN + L2	0.960

Table 1: AUROC scores using CCLR for both FashionMNIST/MNIST and CIFAR10/SVHN dataset pairs using various combinations of model-specific modifications that were proposed in Sect. 2.3. The results included here are the highest AUROC scores from a range of CCLR values calculated using the same k/T-value range as Table 1 in Sect. 3.3. The highest AUROC scores for each dataset pair are in bold.

Here, we present ablation results that support model-specific modifications when using DDPMs for OOD detection. Using L1 rather than L2 as the loss for training and evaluation for both dataset pairs leads to improved AUROC scores. As stated in Sect. 2.3, using the L1 norm has been shown to learn a tighter distribution over the training data and reduce the possibility of hallucinations in DDPMs. we also show that using a linear importance weight, $w(t)$, improves OOD detection performance for both dataset pairs. However, this performance improvement is far more significant in the FashionMNIST/MNIST dataset pair. For the CIFAR10/SVHN dataset pairs, the improvement when using $w(t)$ is minimal when using L1 loss and actually impairs performance when training and evaluating using L2.

B. More Results

Dataset Pair	AUROC scores					
	$\mathcal{L}_\theta^{<T}$	$CCLR_{1/2}$	$CCLR_{1/3}$	$CCLR_{1/5}$	$CCLR_{1/10}$	$CCLR_{1/20}$
MNIST/FashionMNIST	**1.000**	**1.000**	0.996	0.995	0.999	0.999
SVHN/CIFAR10	0.994	0.990	0.994	0.996	**0.997**	0.984

Table 2: AUROC scores for lower-complexity ID dataset pairs MNIST/FashionMNIST and SVHN/CIFAR10. Results are reported for the whole ELBO objective as an OOD score, as well as likelihood ratios for various k/T-values. The highest AUROC scores for each dataset pair are in bold.

In Table 2, we present AUROC results for MNIST/FashionMNIST and SVHN/CIFAR 10 dataset pairs. For both of these comparisons, the less-complex MNIST and SVHN datasets are ID. For each dataset pair, we present AUROC scores using just the ELBO as an OOD score, and for CCLR scores using a range of k/T values. The results show that when a high-complexity dataset is OOD and doesn't corrupt the ELBO term, the ELBO is a strong OOD detection score for both dataset pairs. These results show that the strong performance of the CCLR as an OOD detection score also extends to when less-complex datasets are ID.

C. ROC Curves

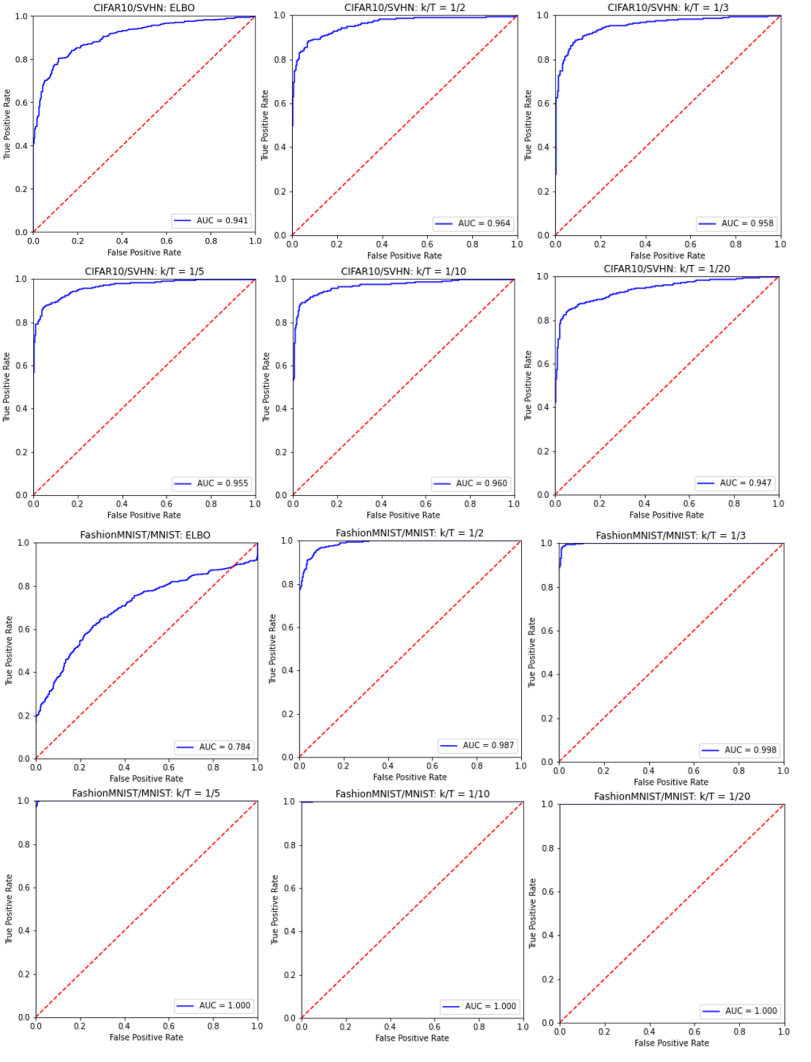

Figure 1: ROC Curves for each experiment in Table 1. ROC Curves are presented using CCLR for various k/T-values as well as the entire ELBO objective directly for both CIFAR10/SVHN and FashionMNIST/MNIST dataset pairs.