A. Supplementary appendix

A.1. Computation time

The amount of overhead depends on the number of auxiliary samples, and the size of the input images. The increase is mainly in the backward pass. With default parameters, our method runs on average at a speed of 67% compared to bilinear sampling for the forward pass, and 40% for the backward pass.

A.2. Results on MNIST

As MNIST dataset is textureless, both sampling methods perform similarly – we used the implementation of [23] and achieved 1.4% error rate for both bilinear sampling and ours. While the two perform similar, as shown on the average aligned image below, ours zooms in more on digits.

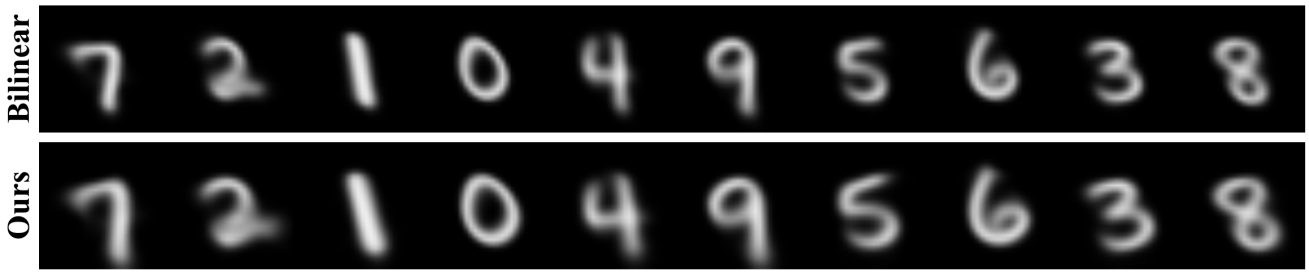

Figure 11: Average aligned image from the STN using (top) bilinear sampling and (bottom) our method.