ViT model, SVHN dataset, $\mathcal{T}$ under WGN

SVHN [1] (http://ufldl.stanford.edu/housenumbers/) is an MNIST-like dataset, but larger, real-world and more challenging.

[1] Netzer, Yuval, et al. "Reading digits in natural images with unsupervised feature learning." (2011).