MedShift: Implicit Conditional Transport for X-Ray Domain Adaptation

Supplementary Material

The supplementary material is organized as follows: Appendix A describes the implementation details of MedShift. Appendix B contains empiric proof of the shared manifold assumption of Section 3.

A. Implementation Details

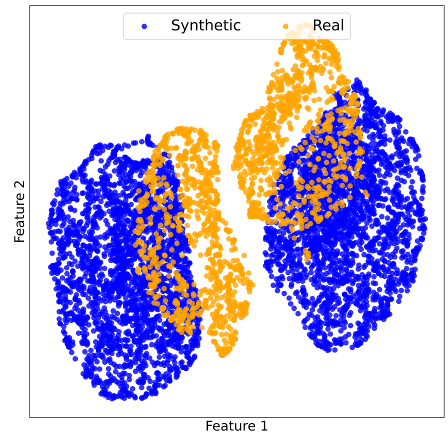
The model was trained on a workstation equipped with an NVIDIA RTX 3090 Ti GPU (24GB VRAM), an Intel Xeon Silver 4216 CPU (2.10 GHz), and 192GB of RAM. We used mixed-precision training via the `Accelerate` [9] library to reduce memory consumption without compromising performance. The hyperparameters used to train MedShift are summarized in Table 7.

Table 7. Model and training configuration used in our experiments.

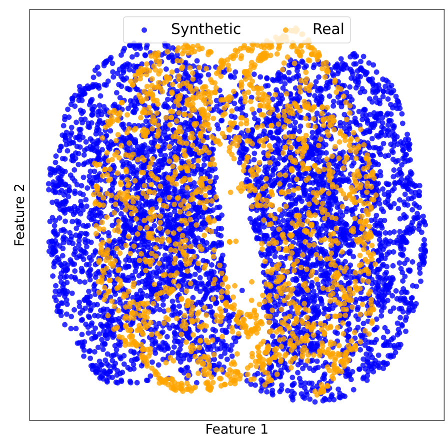
Parameter	Value
Input size	512
Model channels	256
Number of residual blocks	2
Channel multiplier	1, 2, 2, 2
Attention resolutions	2, 4
Number of attention heads	4
Head channels	64
Label Dropout probability	0.2
Learning rate	1e-4
Number of epochs	1,000
Batch size	24
Warmup steps	100
EMA rate	0.999

B. Latent Distributions

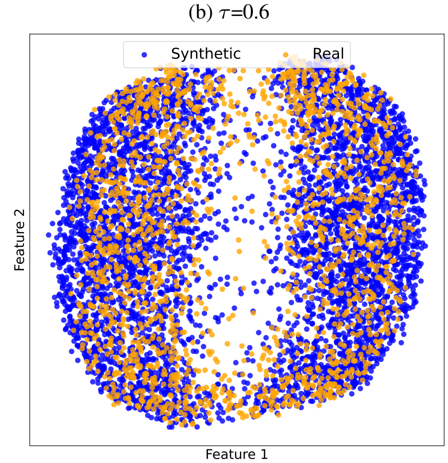
To directly address the assumption of a shared manifold between synthetic and real domains, we add a UMAP analysis of the latent encodings for different τ values in Figure 4. At τ=1.0, where no noise is applied, the embeddings of synthetic and real images remain clearly separated. However, as τ decreases and the model integrates backward, the latent representations become progressively noisier and the two distributions begin to overlap. This analysis supports the core design of our method: by moving to an intermediate, noise-conditioned state, the model converges toward the shared latent space hypothesized in Section 3, enabling the subsequent domain translation.



(a) Original Latents (τ=1.0)



(b) τ=0.6



(c) τ=0.3

Figure 4. UMAP visualization of the latent-space features for different τ levels.