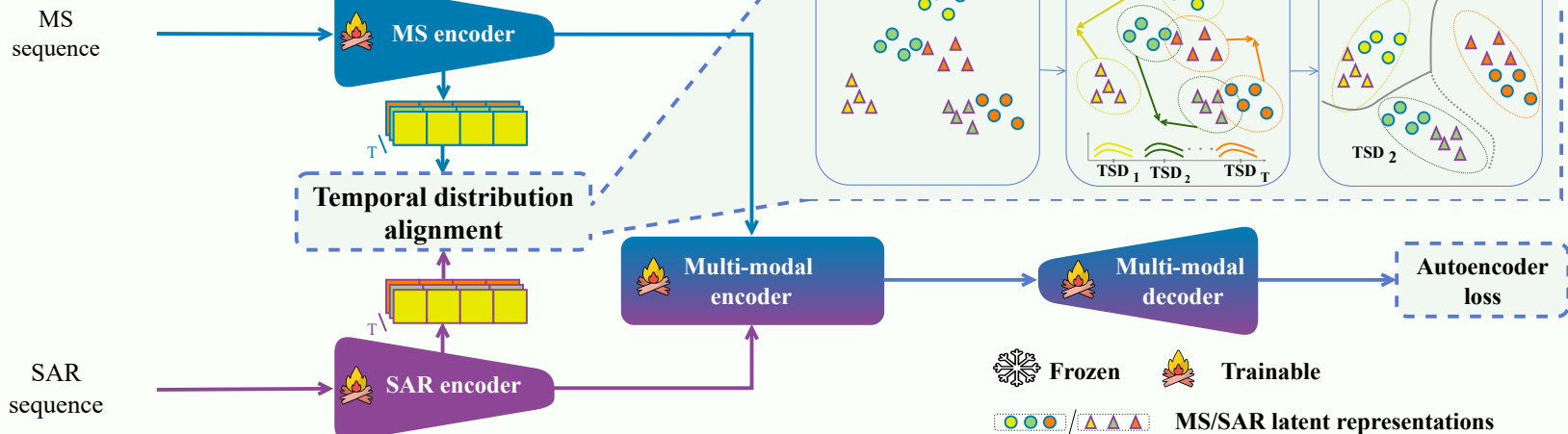


Pre-training phase 1: Temporal Multi-modal Learning



Pre-training phase 2: Latent Reconstruction Learning

