MINT-RVAE: Multi-Cues Intention Prediction of Human-Robot Interaction using Human Pose and Emotion Information from RGB-only Camera Data - APPENDIX

Farida Mohsen[1], Ali Safa[1]

***Abstract*— This is the Appendix to the paper: MINT-RVAE: Multi-Cues Intention Prediction of Human-Robot Interaction using Human Pose and Emotion Information from RGB-only Camera Data.**

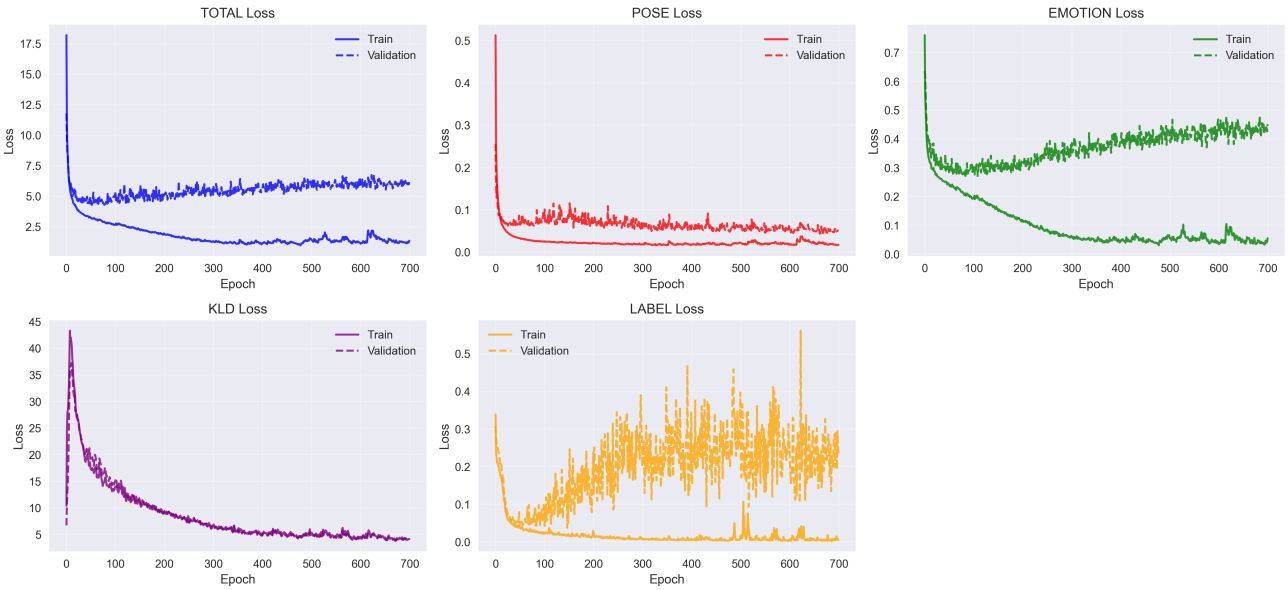

Fig. A1: Training and validation losses for the RVAE across 700 epochs. Top row: total, pose, and emotion reconstruction losses. Bottom row: KL divergence and label reconstruction. Loss stabilization indicates effective multimodal reconstruction with balanced KL regularization.

[1] College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

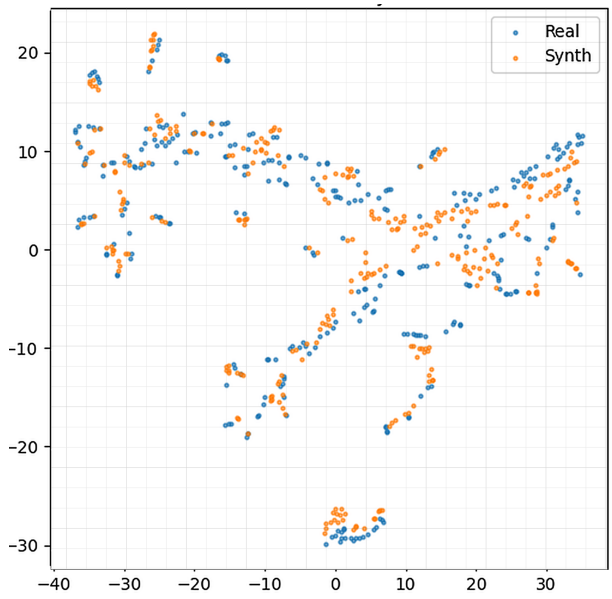

Fig. A2: ***t-SNE visualization of real multimodal data sequences (Real) vs. synthetic sequence embeddings (Synth) produced by our proposed MINT-RVAE model.*** *The significant visual overlap indicates that MINT-RVAE samples correctly cover the training distribution and are able to faithfully model multimodal pose and emotion sequences during synthetic data generation.*