

Train Beta KL-divergence ($B = 1$) loss

