3.2 Variational objective

We train the encoder $q_\phi(z \mid x)$ and the decoder $p_\theta(x \mid z)$ jointly by maximising the evidence lower bound on the marginal log-likelihood of each observation x:

$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{q_\phi(z \mid x)}\big[\log p_\theta(x \mid z)\big] - D_{\mathrm{KL}}\big(q_\phi(z \mid x) \,\|\, p(z)\big). \tag{4}$$

The first term rewards faithful reconstruction; the second keeps the approximate posterior close to the prior $p(z) = \mathcal{N}(0, I)$. Table 2 reports the error for four latent sizes.

Table 2: Mean squared reconstruction error on the held-out split, in thousandths.

Latent size	Baseline	Ours	Change
16	9.84	8.71	−1.13
32	7.62	6.05	−1.57
64	6.48	4.93	−1.55
128	6.31	4.88	−1.43

Increasing d beyond 64 yields diminishing returns, which is consistent with the capacity analysis in Section 5. We therefore fix $d = 64$ in all remaining experiments.