$\tilde{\boldsymbol{y}}$

$\boldsymbol{y}$

$$P(\tilde{\boldsymbol{y}}|\boldsymbol{x}) = \tilde{\boldsymbol{h}}^{(5)}$$

$$P(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{h}^{(5)}$$

$$\tilde{\boldsymbol{h}}^{(5)} = ACT\left(\boldsymbol{\gamma}^{(5)} \odot \left(\tilde{\boldsymbol{z}}^{(5)} + \boldsymbol{\beta}^{(5)}\right)\right)$$

$$\tilde{\boldsymbol{z}}^{(5)} = N_B\left(\boldsymbol{W}^{(5)} \cdot \tilde{\boldsymbol{h}}^{(4)}\right) + noise$$

$$\boldsymbol{u}^{(5)} = N_B\left(\tilde{\boldsymbol{h}}^{(5)}\right)$$
$$\hat{\boldsymbol{z}}_1^{(5)} = g\left(\tilde{\boldsymbol{z}}^{(5)}, \boldsymbol{u}^{(5)}\right)$$
$$\hat{\boldsymbol{z}}_{i,BN}^{(5)} = \frac{\hat{\boldsymbol{z}}_i^{(5)} - \mu_i^{(5)}}{\sigma_i^{(5)}}$$

$C_d^{(5)}$

$$\boldsymbol{h}^{(5)} = ACT\left(\boldsymbol{\gamma}^{(5)} \odot \left(\boldsymbol{z}^{(5)} + \boldsymbol{\beta}^{(5)}\right)\right)$$

$$\boldsymbol{z}_{pre}^{(5)} = \boldsymbol{W}^{(5)} \cdot \boldsymbol{h}^{(4)}$$
$$\boldsymbol{z}^{(5)} = N_B\left(\boldsymbol{z}_{pre}^{(5)}\right)$$

$$\tilde{\boldsymbol{h}}^{(4)} = ACT\left(\boldsymbol{\gamma}^{(4)} \odot \left(\tilde{\boldsymbol{z}}^{(4)} + \boldsymbol{\beta}^{(4)}\right)\right)$$

$$\tilde{\boldsymbol{z}}^{(4)} = N_B\left(\boldsymbol{W}^{(4)} \cdot \tilde{\boldsymbol{h}}^{(3)}\right) + noise$$

$$\boldsymbol{u}^{(4)} = N_B\left(\boldsymbol{V}^{(5)} \cdot \hat{\boldsymbol{z}}^{(5)}\right)$$
$$\hat{\boldsymbol{z}}_1^{(4)} = g\left(\tilde{\boldsymbol{z}}^{(4)}, \boldsymbol{u}^{(4)}\right)$$
$$\hat{\boldsymbol{z}}_{i,BN}^{(4)} = \frac{\hat{\boldsymbol{z}}_i^{(4)} - \mu_i^{(4)}}{\sigma_i^{(4)}}$$

$C_d^{(4)}$

$$\boldsymbol{h}^{(4)} = ACT\left(\boldsymbol{\gamma}^{(4)} \odot \left(\boldsymbol{z}^{(4)} + \boldsymbol{\beta}^{(4)}\right)\right)$$

$$\boldsymbol{z}_{pre}^{(4)} = \boldsymbol{W}^{(4)} \cdot \boldsymbol{h}^{(3)}$$
$$\boldsymbol{z}^{(4)} = N_B\left(\boldsymbol{z}_{pre}^{(4)}\right)$$

$$\tilde{\boldsymbol{h}}^{(3)} = ACT\left(\boldsymbol{\gamma}^{(3)} \odot \left(\tilde{\boldsymbol{z}}^{(3)} + \boldsymbol{\beta}^{(3)}\right)\right)$$

$$\tilde{\boldsymbol{z}}^{(3)} = N_B\left(\boldsymbol{W}^{(3)} \cdot \tilde{\boldsymbol{h}}^{(2)}\right) + noise$$

$$\boldsymbol{u}^{(3)} = N_B\left(\boldsymbol{V}^{(4)} \cdot \hat{\boldsymbol{z}}^{(4)}\right)$$
$$\hat{\boldsymbol{z}}_1^{(3)} = g\left(\tilde{\boldsymbol{z}}^{(3)}, \boldsymbol{u}^{(3)}\right)$$
$$\hat{\boldsymbol{z}}_{i,BN}^{(3)} = \frac{\hat{\boldsymbol{z}}_i^{(3)} - \mu_i^{(3)}}{\sigma_i^{(3)}}$$

$C_d^{(3)}$

$$\boldsymbol{h}^{(3)} = ACT\left(\boldsymbol{\gamma}^{(3)} \odot \left(\boldsymbol{z}^{(3)} + \boldsymbol{\beta}^{(3)}\right)\right)$$

$$\boldsymbol{z}_{pre}^{(3)} = \boldsymbol{W}^{(3)} \cdot \boldsymbol{h}^{(2)}$$
$$\boldsymbol{z}^{(3)} = N_B\left(\boldsymbol{z}_{pre}^{(3)}\right)$$

$$\tilde{\boldsymbol{h}}^{(2)} = ACT\left(\boldsymbol{\gamma}^{(2)} \odot \left(\tilde{\boldsymbol{z}}^{(2)} + \boldsymbol{\beta}^{(2)}\right)\right)$$

$$\tilde{\boldsymbol{z}}^{(2)} = N_B\left(\boldsymbol{W}^{(2)} \cdot \tilde{\boldsymbol{h}}^{(1)}\right) + noise$$

$$\boldsymbol{u}^{(2)} = N_B\left(\boldsymbol{V}^{(3)} \cdot \hat{\boldsymbol{z}}^{(3)}\right)$$
$$\hat{\boldsymbol{z}}_1^{(2)} = g\left(\tilde{\boldsymbol{z}}^{(2)}, \boldsymbol{u}^{(2)}\right)$$
$$\hat{\boldsymbol{z}}_{i,BN}^{(2)} = \frac{\hat{\boldsymbol{z}}_i^{(2)} - \mu_i^{(2)}}{\sigma_i^{(2)}}$$

$C_d^{(2)}$

$$\boldsymbol{h}^{(2)} = ACT\left(\boldsymbol{\gamma}^{(2)} \odot \left(\boldsymbol{z}(2) + \boldsymbol{\beta}^{(2)}\right)\right)$$

$$\boldsymbol{z}_{pre}^{(2)} = \boldsymbol{W}^{(2)} \cdot \boldsymbol{h}^{(1)}$$
$$\boldsymbol{z}^{(2)} = N_B\left(\boldsymbol{z}_{pre}^{(2)}\right)$$

$$\tilde{\boldsymbol{h}}^{(1)} = ACT\left(\boldsymbol{\gamma}^{(1)} \odot \left(\tilde{\boldsymbol{z}}^{(1)} + \boldsymbol{\beta}^{(1)}\right)\right)$$

$$\tilde{\boldsymbol{z}}^{(1)} = N_B(\boldsymbol{W}^{(1)} \cdot \tilde{\boldsymbol{h}}^{(0)}) + noise$$

$$\boldsymbol{u}^{(1)} = N_B\left(\boldsymbol{V}^{(2)} \cdot \hat{\boldsymbol{z}}^{(2)}\right)$$
$$\hat{\boldsymbol{z}}_1^{(1)} = g\left(\tilde{\boldsymbol{z}}^{(1)}, \boldsymbol{u}^{(1)}\right)$$
$$\hat{\boldsymbol{z}}_{i,BN}^{(1)} = \frac{\hat{\boldsymbol{z}}_i^{(1)} - \mu_i^{(1)}}{\sigma_i^{(1)}}$$

$C_d^{(1)}$

$$\boldsymbol{h}^{(1)} = ACT\left(\boldsymbol{\gamma}^{(1)} \odot \left(\boldsymbol{z}^{(1)} + \boldsymbol{\beta}^{(1)}\right)\right)$$

$$\boldsymbol{z}_{pre}^{(1)} = \boldsymbol{W}^{(1)} \cdot \boldsymbol{h}^{(0)}$$
$$\boldsymbol{z}^{(1)} = N_B\left(\boldsymbol{z}_{pre}^{(1)}\right)$$

$$\tilde{\boldsymbol{h}}^{(0)} = \tilde{\boldsymbol{z}}^{(0)}$$

$$\tilde{\boldsymbol{z}}^{(0)} = \boldsymbol{x} + noise$$

$$\boldsymbol{u}^{(0)} = N_B\left(\boldsymbol{V}^{(1)} \cdot \hat{\boldsymbol{z}}^{(1)}\right)$$
$$\hat{\boldsymbol{z}}_0^{(0)} = g\left(\tilde{\boldsymbol{z}}^{(0)}, \boldsymbol{u}^{(0)}\right)$$
$$\hat{\boldsymbol{z}}_{i,BN}^{(0)} = \frac{\hat{\boldsymbol{z}}_i^{(0)} - \mu_i^{(0)}}{\sigma_i^{(0)}}$$

$C_d^{(0)}$

$$\tilde{\boldsymbol{h}}^{(0)} = \boldsymbol{z}^{(0)}$$

$$\boldsymbol{z}^{(0)} = \boldsymbol{x} + noise$$

x

Corrupted Encoder   Decoder   Clean Encoder

$$\boldsymbol{\mu}^{(l)} = batchmean\left(\boldsymbol{z}_{pre}^{(l)}\right)$$

$$\boldsymbol{\sigma}^{(l)} = batchstd\left(\boldsymbol{z}_{pre}^{(l)}\right)$$

$$\hat{z}_i^{(l)} = g_i\left(\tilde{z}_i^{(l)}, u_i^{(l)}\right) = \left(\tilde{z}_i^{(l)} - \mu_i\left(u_i^{(l)}\right)\right) \cdot v_i\left(u_i^{(l)}\right) + \mu_i\left(u_i^{(l)}\right)$$

$$\mu_i\left(u_i^{(l)}\right) = a_{1,i}^{(l)} \cdot sigmoid\left(a_{2,i}^{(l)} \cdot u_i^{(l)} + a_{3,i}^{(l)}\right) + a_{4,i}^{(l)} \cdot u_i^{(l)} + a_{5,i}^{(l)}$$

$$v_i\left(u_i^{(l)}\right) = a_{6,i}^{(l)} \cdot sigmoid\left(a_{7,i}^{(l)} \cdot u_i^{(l)} + a_{8,i}^{(l)}\right) + a_{9,i}^{(l)} \cdot u_i^{(l)} + a_{10,i}^{(l)}$$

$$C_d = \sum_{l=0}^{L} C_d^{(l)} = \sum_{l=0}^{L} ||\boldsymbol{z}^{(l)} - \hat{\boldsymbol{z}}_{BN}^{(l)}||^2$$