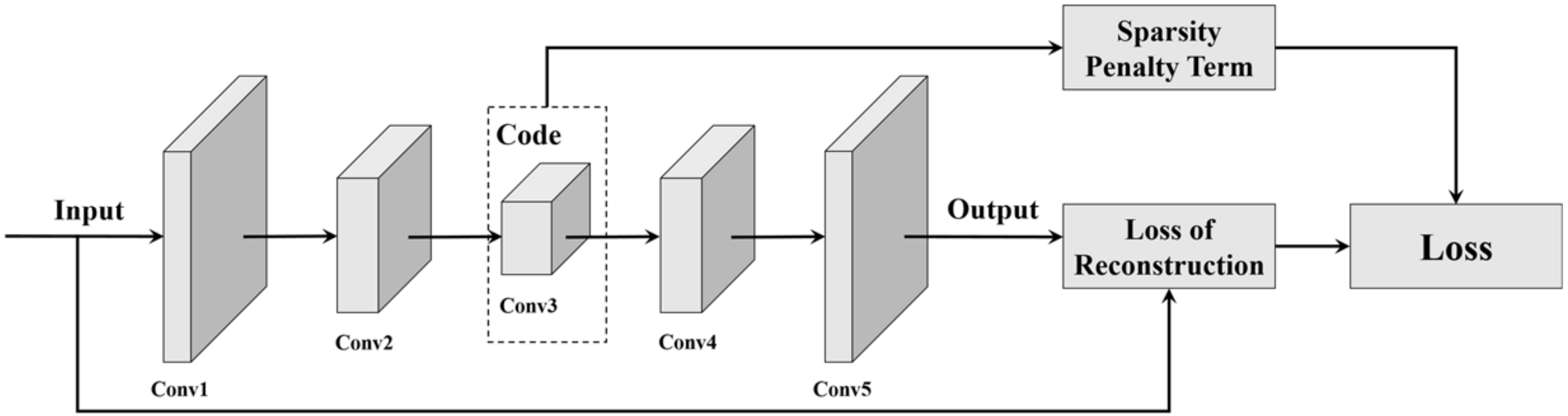


Sparse Autoencoder Loss

$$J_{sparse} = J_{mse} + \beta \cdot \sum_{j=1}^{s_{code}} KL(\rho || \hat{\rho}_j)$$

$$J_{mse} = \frac{1}{2M} \sum_{m}^{M} (\hat{y}_m - y_m)^2$$

$$KL(\rho || \hat{\rho}_j) = \rho \cdot log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \cdot log \frac{1 - \rho}{1 - \hat{\rho}_j}$$

$$\hat{\rho}_j = \frac{1}{n} \sum_{i}^{n} a_j^{(code)}$$

β: weight of sparsity penalty term

s_{code}: the number of hidden units in layer “code”

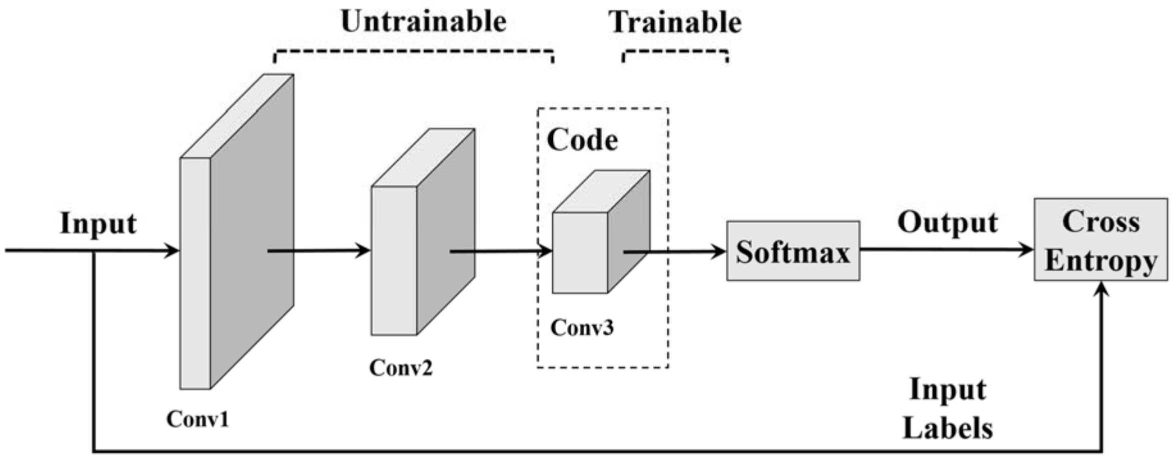
$\hat{y}_m$: reconstructed voxel’s intensity

y_m: original voxel’s intensity

ρ: sparsity parameter, manually set

$\hat{\rho}_j$: average activation of filter j

$a_j^{(code)}$: the activation of jth hidden unit in layer “code”


Untrainable
Trainable
Code
Input
Softmax
Output
Cross Entropy
Conv1
Conv2
Conv3
Input Labels