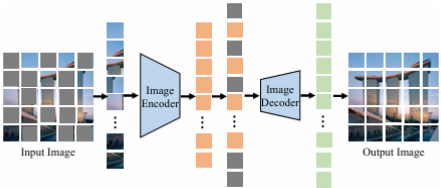
Masked Image Modeling - Generative



Randomly masks a subset of image patches and trains the model to reconstruct them from the unmasked patches.

Loss Formulation

$$\mathcal{L}_{\text{MIM}} = -\frac{1}{B}\sum_{i=1}^{B} \log f_\theta(x_i^I \mid \hat{x}_i^I)$$

where x_i^I is the masked patch set and $\hat{x}_i^I$ is the unmasked patch set of image x_i^I.