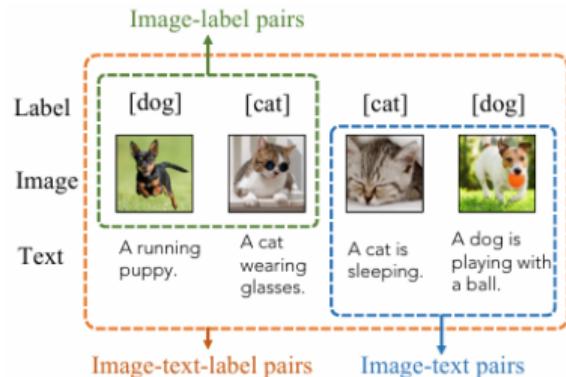


Image-Text-Label Contrastive Learning



Introduces supervised contrastive learning into image-text contrastive learning by grouping samples with the same label as positives.

Loss Formulation

$$\mathcal{L}_{I \rightarrow T}^{\text{ITL}} = - \sum_{i=1}^B \frac{1}{|\mathcal{P}(i)|} \sum_{k \in \mathcal{P}(i)} \log \frac{\exp(z_i^I \cdot z_k^T / \tau)}{\sum_{j=1}^B \exp(z_i^I \cdot z_j^T / \tau)}$$

$$\mathcal{L}_{T \rightarrow I}^{\text{ITL}} = - \sum_{i=1}^B \frac{1}{|\mathcal{P}(i)|} \sum_{k \in \mathcal{P}(i)} \log \frac{\exp(z_i^T \cdot z_k^I / \tau)}{\sum_{j=1}^B \exp(z_i^T \cdot z_j^I / \tau)}$$