Improving MB

(P) Improving MB

given images $x_1 \cdots x_n$ & their encoding $f(x_i) = \phi(x_i, \theta) \in \mathbb{R}^d$ (NN)

$$\mathcal{L}_{CE} = \log \prod_{i=1}^{n} P(i|x_i) = \sum_{i=1}^{n} \log \frac{e^{(f_i^T f_i/\tau)}}{\sum_{j=1}^{n} \exp(f_i^T f_j/\tau)} \quad (I)$$ Normal

mini large batch

K augmentation for each $x_m \rightarrow \{x_m^{(1)}, x_m^{(2)} \cdots x_m^{(k)}\}$

$$\text{modified } \mathcal{L}_{CE} = \sum_{i=1}^{|B|} \sum_{k=1}^{K} \log \frac{\exp(f_i^T f_i^{(k)}/\tau)}{\sum_{j=1}^{n} \exp(\tilde{f}_j f_i^{(k)}/\tau)} \quad (II)$$

(1) two sum; $\tilde{f}_j$: Near embed.

$$\tilde{f}_i = m \tilde{f}_i + (1-m) \sum_{k=1}^{K} \frac{1}{K} f_i^{(k)}$$ // Aggregation with some moment.

Instance consistency (2)

eq (I)

$$\text{Consistency loss, } \mathcal{L}_{cons} = \sum_{k=1}^{K} \sum_{j \neq k} KL\left(P(i|x_i^{(k)}) \,||\, P(i|x_j^{(j)})\right) \quad (III)$$

$x_i^{(k)}$ – $x_j^{(j)}$: diff; $x_i^{(k)}$ – $x_i^{(j)}$: same

the whole thing is contrastive

classified different [Diffent embedding looks different]