①

① what should be contrastive

What should be contrastive

view Generations:

n atomic augmentation.

query view I_q

first key view $I_{k_0} := \{q, k_0\} = T\{x_1^{\{q,k_0\}}, x_2^{\{q,k_0\}}, \ldots, x_n^{\{q,k_n\}}\}$

n views from reference images $I_{k_i}, \forall i \in \{1, \ldots, n\}$

Provides with I_q, I_{k_0}, I_{k_1} ...

so on so forth.

Contrastive Embedding Space:

$f(): \mathcal{X} \rightarrow v \in \mathbb{R}^d$ → 1, 2, ... n: nthead each $\mathbb{R}^d$

MTL setup:

$v^q, v^{k_0} \ldots v^{k_n} \rightarrow \mathbb{R}^d$

projected into (n+1) normalized embedding.

$z_0, z_1 \ldots z_n \in \mathbb{R}^{d'}$ by head $h: v \rightarrow z$

$x \rightarrow q$, k_0, ..., k_n

z_0 ↓ invariant to all (q)

z_1 ↓ dependent on 1st transformation ... so on so forth

(P) what should be (11)

same image

↑ +/− ↷ Different instance

embedding → $z_i^{k_i^{+/-}}$ → Augmentation.

↓

~~instance~~ head

image → $Z_i^{k_j^{+/-}}$ → Aug → with ~~Augmentation~~ head

which instance same (+) diff (−)

Augmentation

three things to keep track

$$E_{ij}^{\{+,-\}} = \exp\left(\frac{z_i^q \cdot z_i^{k_j\{+,-\}}}{\tau}\right)$$

overall loss

$$\mathcal{L}_q = -\frac{1}{n+1}\left[\log \frac{E_{0,0}^{+}}{E_{0,0}^{+} + \sum_{k^-} E_{0,0}^{-}}\right]$$

$$Z_i^{k_j^{+/-}}$$

$$-\frac{1}{n+1}\left[\log \frac{E_{i,i}^{+}}{\sum_{j=0}^{n} E_{i,j}^{+} + \sum_{k^-} E_{i,i}^{-}}\right]$$