

① what should be contrastive

What should be contrastive

view generations:

$n$  atomic augmentation.

query view  $I_q$

first key view  $I_{k_0} \approx \{q, k\} = \mathcal{T} \{x_1^{q, k_0}, x_2^{q, k_0}, \dots, x_n^{q, k_0}\}$

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

Provides with  $I_q, I_{k_0}, I_{k_1}, \dots$   
so on so forth.

Contrastive Embedding Space:

$f() : \mathcal{X} \rightarrow v \in \mathbb{R}^d$   $\begin{matrix} \uparrow \\ \text{at least} \end{matrix}$   $\begin{matrix} \downarrow \\ \text{each} \end{matrix}$   $\mathbb{R}^d$

MTE setup:

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

projected into  $(n+1)$  normalized embedding.

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

$x \rightarrow q$   
 $k_0$   
 $\vdots$   
 $k_n$

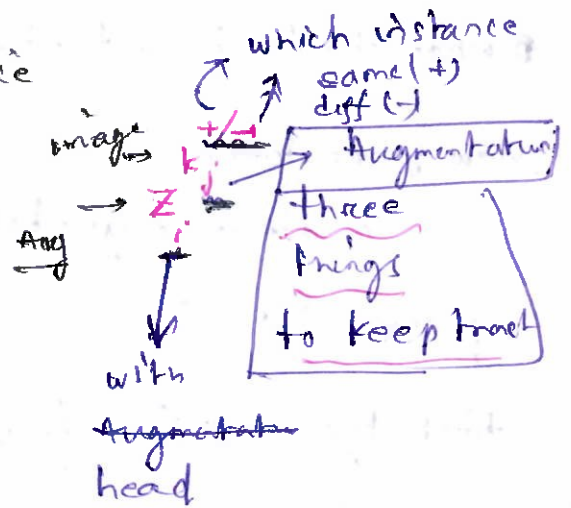
invariant to all  $(q)$

dependent on 1st transformation  $\dots$  so on so forth

⑧ what should be same image ⑪

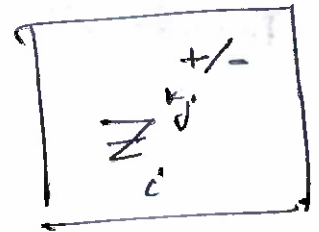
↑ different instance  
 embedding  $\rightarrow z_i$   $\xrightarrow{k_i}$  augmentation  
 ↓ instance head

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



overall loss

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



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