Temporal Cycle-consistency learning

Ⓟ Temporal cycle-consistency learning

Notation: frame Sequence $S = \{s_1, s_2 \cdots s_N\}$ (frame)

Embedding $u_i = \phi(s_i; \theta)$ | image based??

Two video sequences: S, T (N, M length.)

Embedding $\{u_1, u_2 \cdots u_N\}$, $\{v_1, v_2 \cdots v_M\}$

U V

Cycle Consistency: $u_i \in U$

find $v_j = \arg\min_{v \in V} \|u_i - v\|$ // Nearest neighbor of u_i

Repeating $u_k = \arg\min_{u \in U} \|v_j - u\|$ // Nearest Neighbor of v_j

The points are cycle Consistent iff $i \equiv k$

track it to U set.

Cyck back classification:

Soft Nearest Neighbor: $\tilde{v} = \sum_{j=1}^{M} \alpha_j v_j$ (input point)

Softmax weight: $\alpha_j = \frac{e^{-\|u_i - v_j\|^2}}{\sum_{k}^{M} e^{-\|u_i - v_k\|^2}}$

N class classification problem: Logit $x_k = -\|\tilde{v} - u_k\|_2^2$

$\hat{y} = \text{softmax}(x)_k$

{ the smaller the better

Cross - Entropy loss:

only the $u_i \uparrow$ is ↑

$$L_{cbc} = -\sum_{j}^{N} y_i \log(\hat{y}_j)$$

Cycle back - regression:

$\in \cup$

Maximity Similarity vectors $\beta_k = \dfrac{\|\tilde{v} - u_k\|^2}{\sum_{j=1}^{K} \|\tilde{v} - u_k\|^2}$

variance

regularization

peaky around i

Regularizer

$$L_{Cbrc} = \frac{|i - \mu|^2}{\sigma^2} + \lambda \log \sigma$$

where,
$$\mu = \sum_{k=1}^{N} \beta_k k$$
$$\sigma^2 = \sum_{k=1}^{N} \beta_k * (k - \mu)^2$$

The losses are differentiable

Backpropagation