Soft Ranking Threshold Losses for Image Retrieval

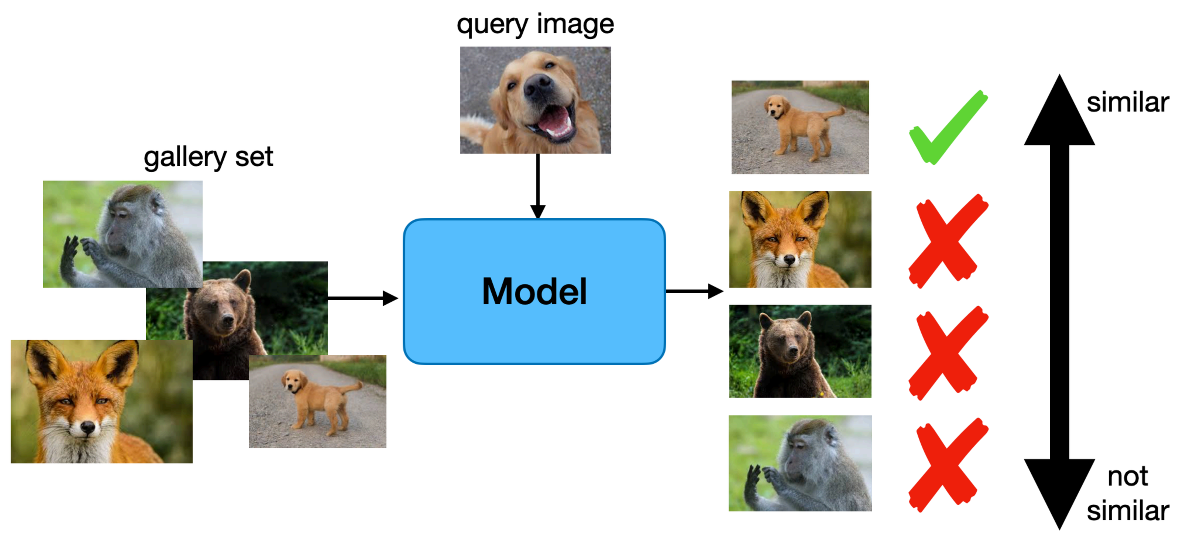
國立臺灣大學 National Taiwan University
國立交通大學 National Chiao Tung University

Chiao-An Yang, Zhixiang Wang, Yen-Yu Lin, Yung-Yu Chuang

ICIP 2021
Paper ID: 2962

Introduction

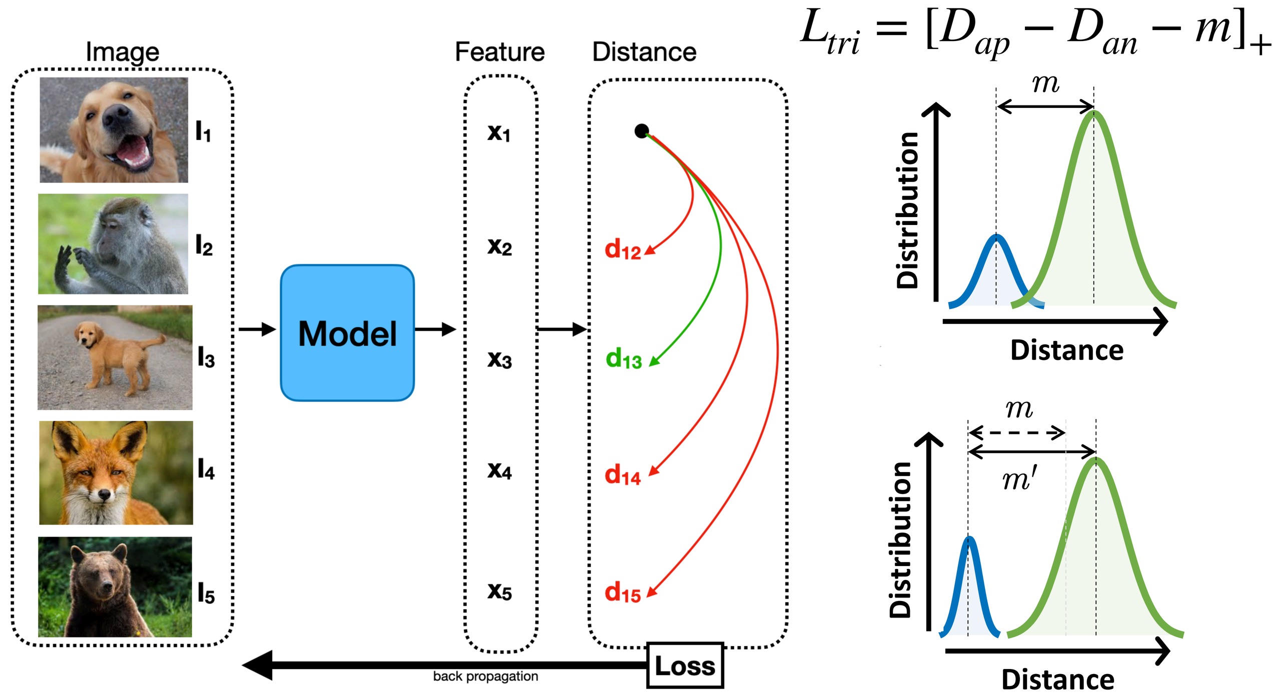
- **Image Retrieval**



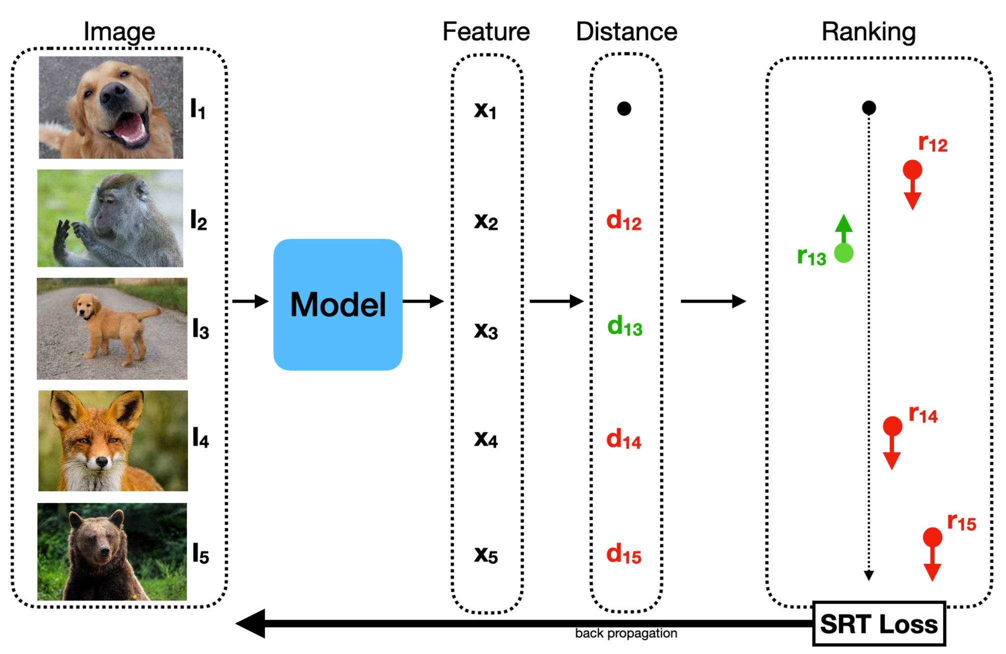
- **Distance-based metric learning**

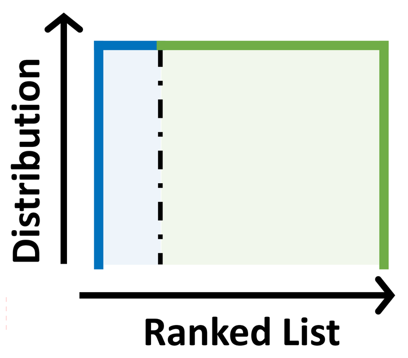
Problem: The distribution of distances varies from batch to batch, making it hard to pre-select optimal hyper-parameters.

$$L_{tri} = [D_{ap} - D_{an} - m]_+$$



- **Ranking-based metric learning**





Ranking

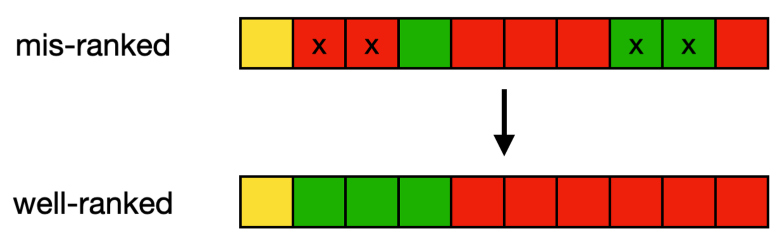
(1) The distribution is more **uniform**.

(2) The rank values have a fixed **lower bound and upper bound**.

(3) A single *rank* value contains **more information** than a single *distance* value.

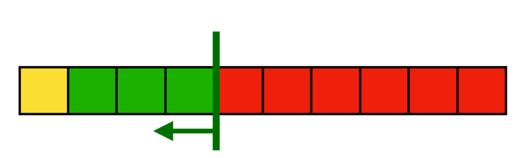
Approach

- **Ranking Threshold Loss**



$$L = \begin{cases} [R_{ij} - T_i^+]_+ & \text{if } y_i = y_j \\ [T_i^- - R_{ij}]_+ & \text{if } y_i \neq y_j \end{cases}$$

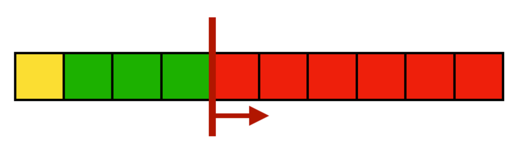
(1) positive term



$$P_i = \#\{j \mid y_i = y_j\}$$

$$T_i^+ = P_i$$

(2) negative term



$$P_i = \#\{j \mid y_i = y_j\}$$

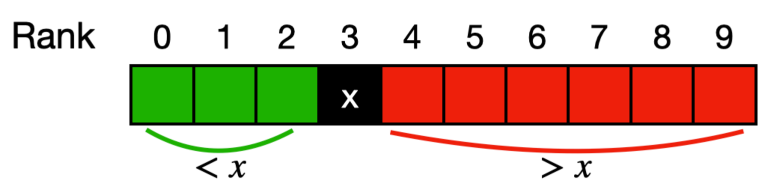
$$T_i^- = P_i + 1$$

(3) combined

$$L_{SRT} = \alpha L_{pos} + (1 - \alpha) L_{neg}$$

- **Soft Ranking**

Problem: Ranking function is not differentiable.

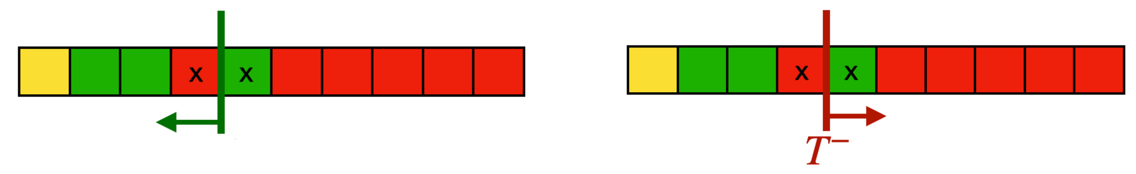


$$\begin{aligned} R(x) &= 3 \\ &= \#\{i \mid i < x\} = \#\{i \mid x - i > 0\} \\ &= \sum_i step(x - i) \\ &\approx \sum_i sigmoid(x - i) = \tilde{R}(x) \end{aligned}$$

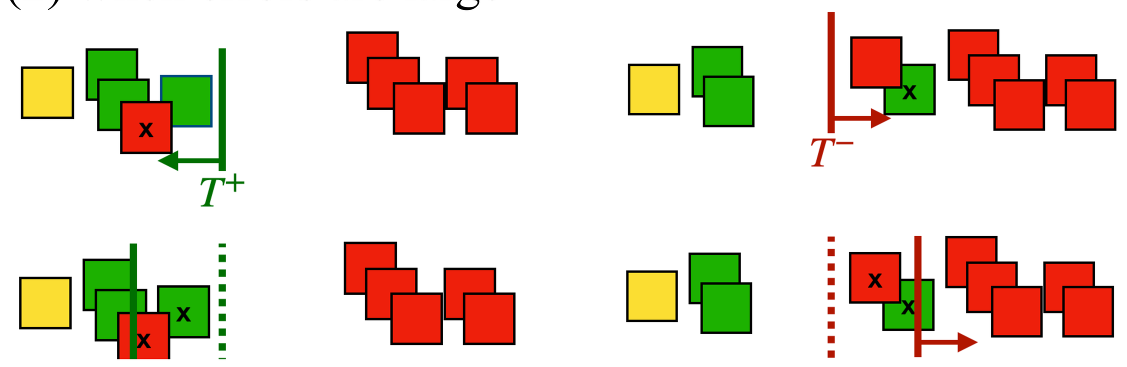
- **Hard Thresholds**

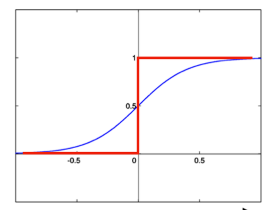
Problem: The approximation errors from soft ranking will accumulate and thus not always neglectable.

(1) when errors are small



(2) when errors are large



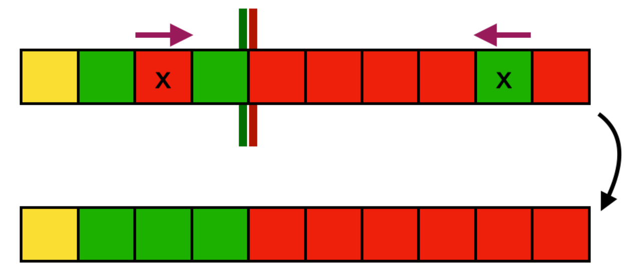


$$\begin{cases} T_i^+ = P_i \\ T_i^- = P_i + 1 \end{cases} \longrightarrow \begin{cases} \tilde{T}_i^+ = \dfrac{P_i}{2} \\ \tilde{T}_i^- = \dfrac{B + P_i + 1}{2} \end{cases}$$

- **Ranking Margins**

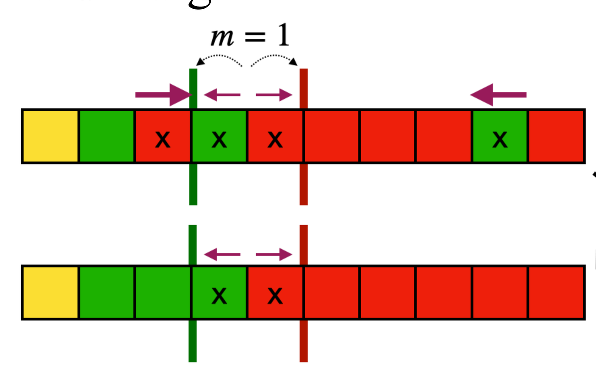
Problem: Since only a few samples contribute to the SRT loss, once these samples are corrected, the training stop.

(1) no margin



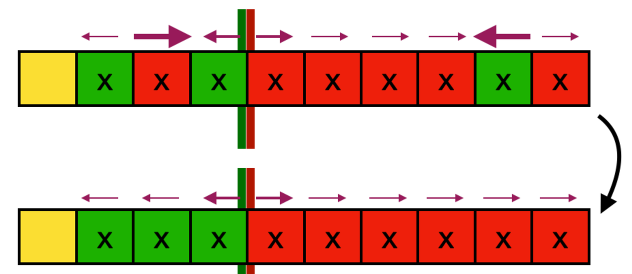
$$L_{pos} = [R_{ij} - T_i^+]_+$$
$$L_{neg} = [T_i^- - R_{ij}]_+$$

(2) hard margin



$$L_{pos} = [R_{ij} - (T_i^+ - m)]_+$$
$$L_{neg} = [(T_i^- + m) - R_{ij}]_+$$

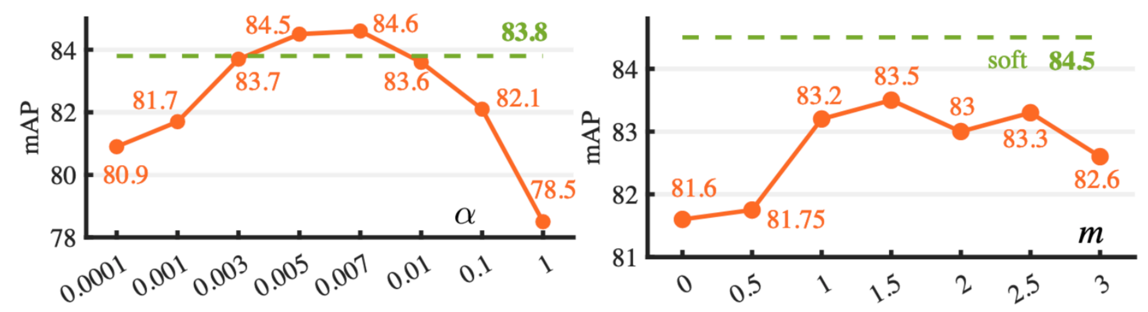
(3) soft margin



$$L_{pos} = \sigma_{sp}(R_{ij} - T_i^+)$$
$$L_{neg} = \sigma_{sp}(T_i^- - R_{ij})$$

Experiments

- **Performance Study**



(a) Study of α (b) Study of m

Basic form	+Hard threshold	+Ranking margin	Full
81.6	81.8	84.5	**84.7**

(c) Ablation study

- **Person ReID**

Loss Type	Market1501		CUHK03-NP detected		CUHK03-NP labeled		DukeMTMC	
	mAP	CMC@1	mAP	CMC@1	mAP	CMC@1	mAP	CMC@1
Softmax	53.8	79.2	23.6	28.9	27.9	31.8	47.0	68.3
Tri	68.7	85.0	52.4	58.9	56.4	62.9	59.1	76.2
Tri-HN	73.0	87.9	55.3	61.4	58.4	65.1	62.9	79.7
Tri-BH	74.0	88.5	56.0	59.4	58.9	64.4	63.6	78.9
Tri-AW	75.3	89.4	58.2	64.0	60.7	66.9	64.1	80.2
Tri-AW*	76.5	89.7	58.9	64.5	61.1	67.1	65.5	81.4
SRT	77.3	90.1	59.4	65.3	62.9	68.6	65.9	81.1
SRT-F	78.6	90.3	62.1	67.5	65.1	70.5	**68.3**	82.3
SRT-F*	**79.2**	**90.8**	**63.0**	**68.1**	**65.8**	**71.3**	67.9	**83.0**

- **Fashion Retrieval**

Loss Type	Consumer-to-shop w/o bbox		Consumer-to-shop w/ bbox		In-shop w/o bbox		In-shop w/ bbox	
	mAP	CMC@20	mAP	CMC@20	mAP	CMC@1	mAP	CMC@1
Softmax	×	×	×	×	53.0	73.2	51.3	71.2
Tri.	13.8	45.0	20.2	57.0	65.4	81.6	65.5	81.8
Tri-HN	20.6	56.5	28.8	68.5	69.2	85.4	68.1	84.4
Tri-BH	×	×	26.3	64.1	69.3	85.2	68.1	83.5
Tri-AW	19.7	55.8	27.1	66.9	70.4	85.9	70.3	85.8
SRT	19.4	55.1	27.3	67.0	71.4	**87.2**	**71.4**	**86.9**
SRT-F	**21.2**	**58.0**	**28.9**	**68.7**	**71.6**	86.9	71.2	86.5

Conclusion

- We propose a novel loss function using the ranking as input, with both a positive and a negative term, to assert the ranking values to satisfy certain adaptive thresholds.
- We introduce the hard thresholds and ranking margin as extensions for further improving its performance.
- Experiments on person reID and fashion retrieval benchmarks demonstrate that our loss outperforms other distance-based losses