

QuRe: Query-Relevant Retrieval through Hard Negative Sampling in Composed Image Retrieval

Jaehyun Kwak, Izaaz Inhar, Se-Young Yun, Sung-Ju Lee

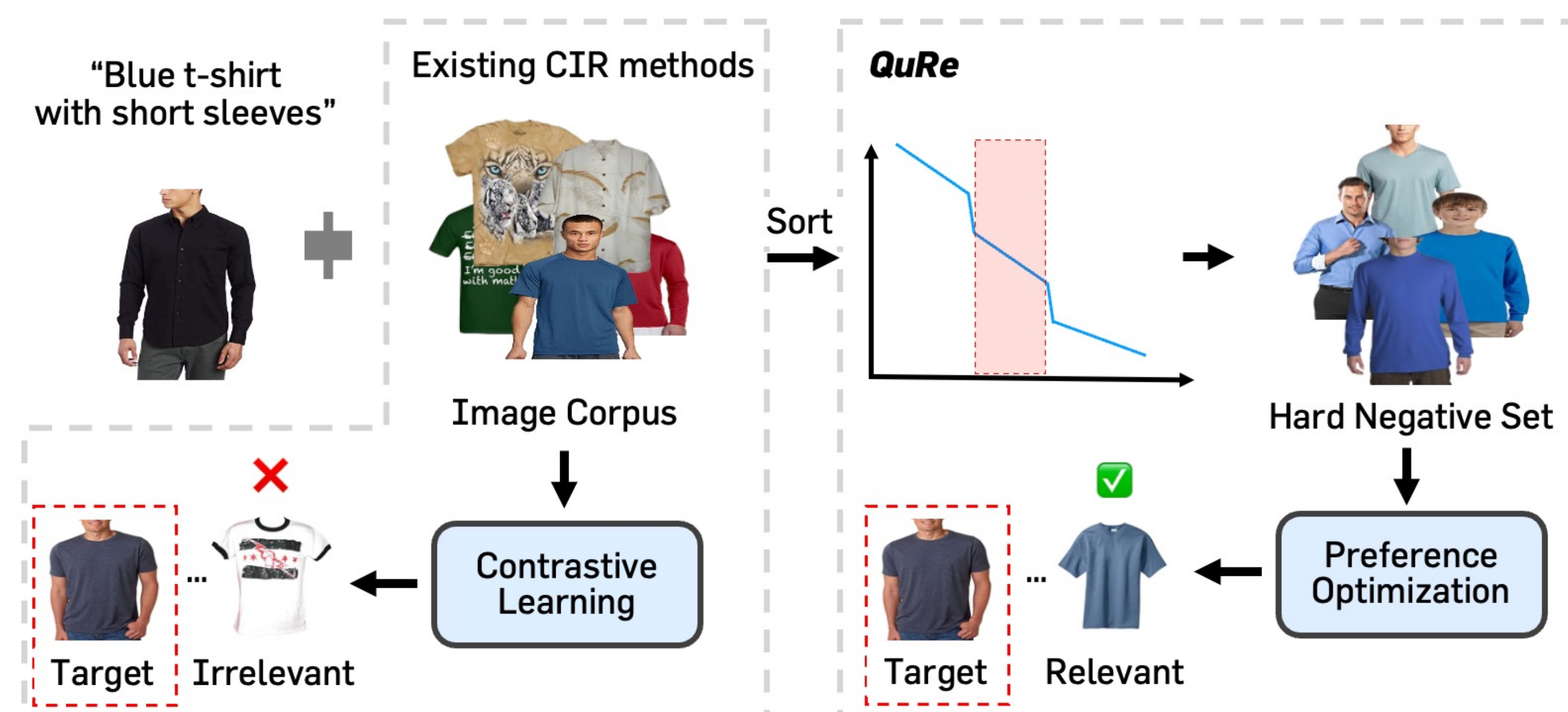
Github



TL;DR

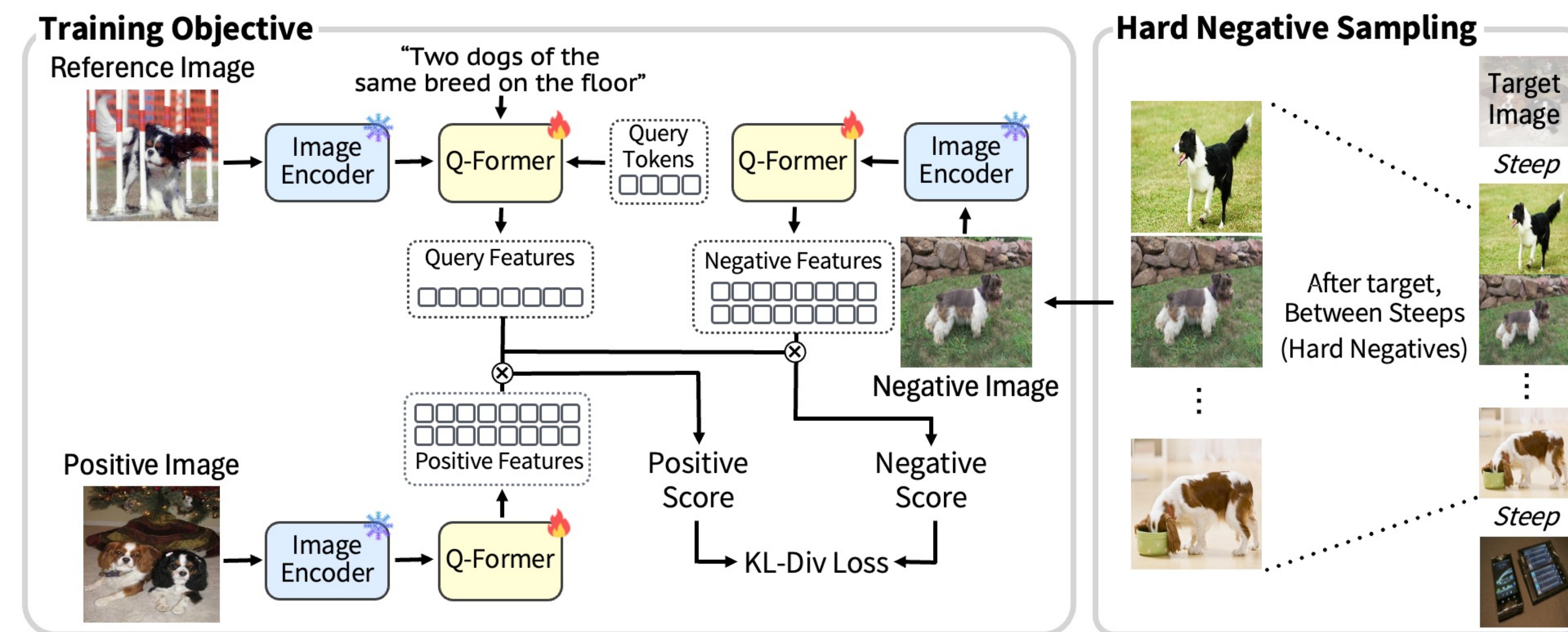
- ✓ **QuRe** : CIR model that ranks both the target and other relevant images at the top
- ✓ **HP-FashionIQ** : Human-preference CIR dataset for assessing how well models align with user judgments

QuRe Overview



- ✓ Existing CIR methods retrieve the target correctly yet still place **irrelevant images among top results**
- ✓ QuRe ranks **both the target and other images highly**

Methodology

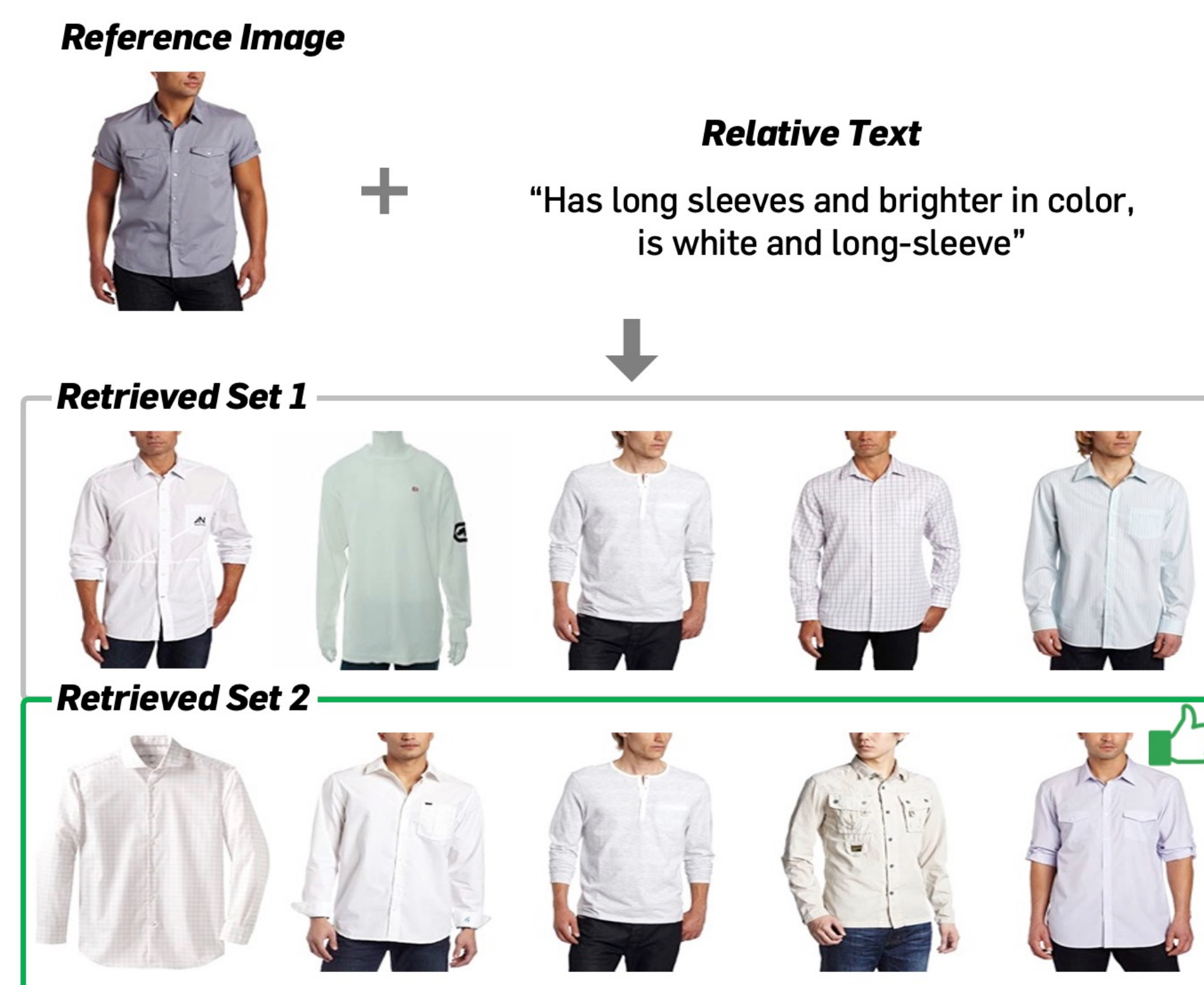


- ✓ **Preference Optimization**
Assign one negative per query, exclude false negatives
- ✓ **Hard-negative Mining**
Update hard negative set by selecting images between the two largest post-target score drops

HP-FashionIQ dataset

- ✓ **HP-FashionIQ** provides human relevance judgements for each retrieved image, annotated by 61 participants on the FashionIQ validation set
- ✓ HP-FashionIQ supports the **Preference Ratio** metric, quantifying how often a model's higher-scoring set aligns with human preference:

$$P(A \succ B \mid s_{rel}(A) > s_{rel}(B))$$



Experiment Results

Hard negative set



FashionIQ / CIRR

Method	Dress		Shirt		Topslee		Average		
	R@10	R@50	R@10	R@50	R@10	R@50	R@10	R@50	Avg.
CoVR-BLIP (Ventura et al., 2024b)	44.55	69.03	48.43	67.42	52.60	74.31	48.53	70.25	60.24
SPRC (Bai et al., 2024)	45.71	70.00	51.37	72.77	55.48	77.46	50.86	73.41	62.13
QuRe	46.80	69.81	53.53	72.87	57.47	77.77	52.60	73.48	63.04

Method	Recall@K				Recall _g @K			Average
	K=1	K=5	K=10	K=50	K=1	K=2	K=3	
CoVR-BLIP (Ventura et al., 2024b)	39.76	70.15	80.89	95.01	72.46	87.86	94.77	71.30
SPRC (Bai et al., 2024)	50.75	80.58	88.72	97.59	79.57	91.76	96.70	80.07
QuRe	52.22	82.53	90.31	98.17	78.51	91.28	96.48	80.52

- ✓ QuRe's hard negatives **lies close to the target but at lower relevance**
- ✓ QuRe achieves the **highest Recall@k** and **best aligns with human preference**

HP-FashionIQ

Method	Preference Rate (%)
CosMo (Lee et al., 2021)	72.96
MGUR (Chen et al., 2024)	73.99
CLIP4CIR (Baldri et al., 2023)	74.45
Bi-BLIP4CIR (Liu et al., 2024)	67.33
CoVR-BLIP (Ventura et al., 2024b)	73.15
SPRC (Bai et al., 2024)	73.82
QuRe	74.55