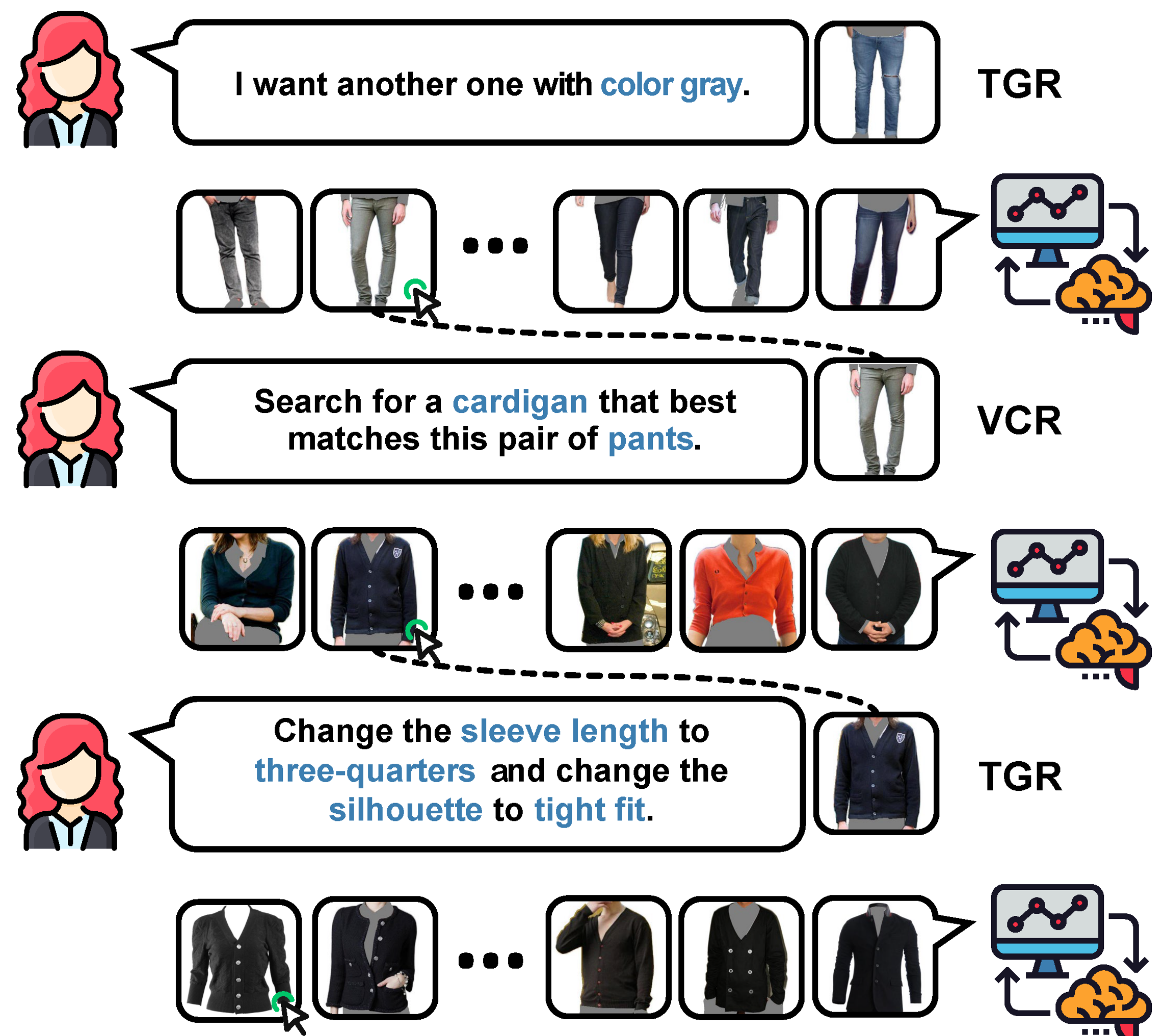


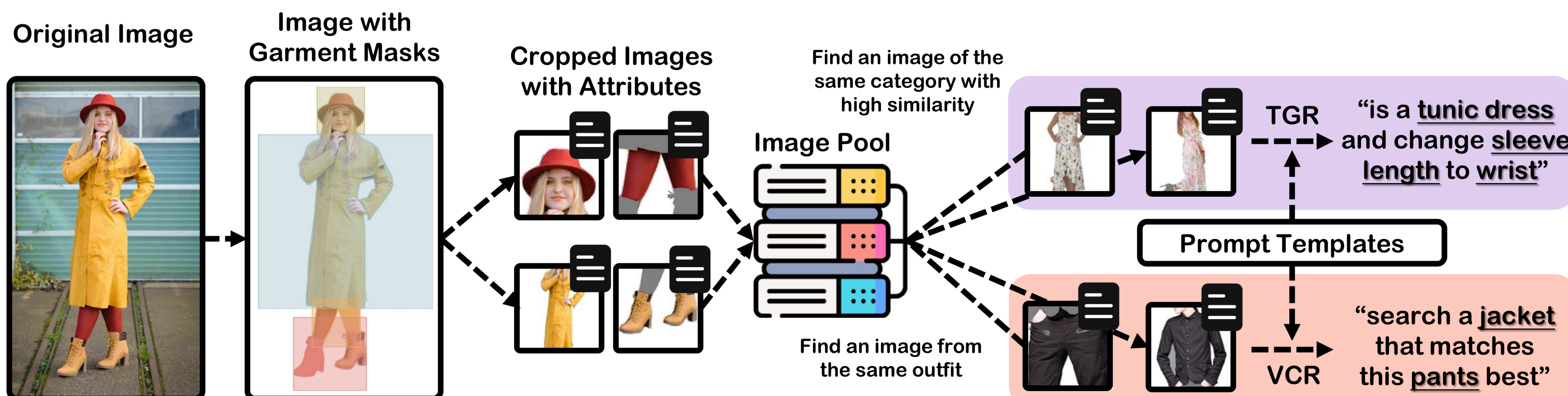
Problem



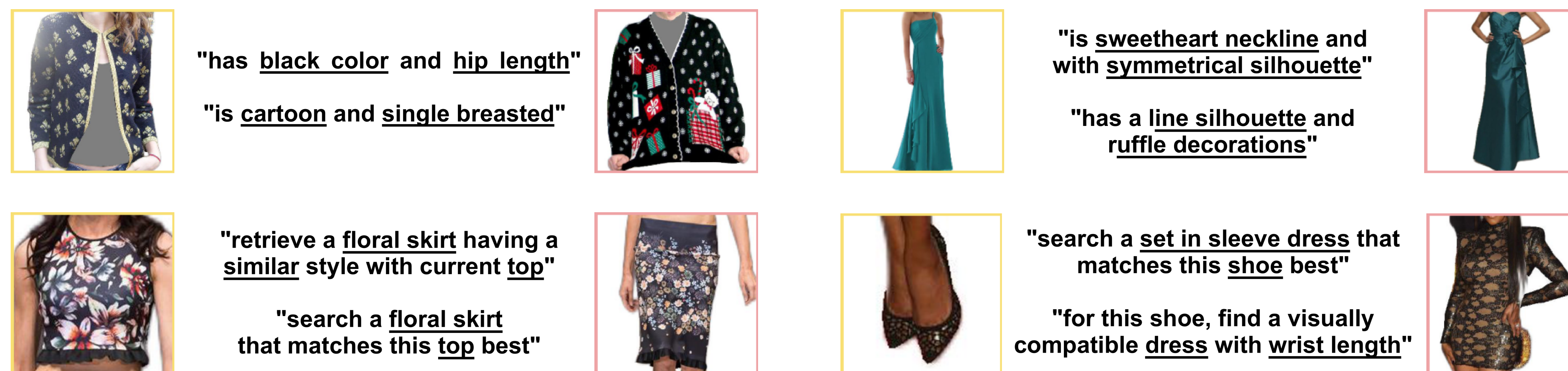
Highlights

- We propose a unified setting for interactive garment retrieval (UIGR), including both Text-Guided garment Retrieval (TGR) and Visually Compatible garment Retrieval (VCR).
- We establish a novel benchmark for the study of this unified problem by re-purposing Fashionpedia.
- We introduce a multi-task model jointly learning two tasks and serves as a strong baseline for UIGR.

Data Collection Process



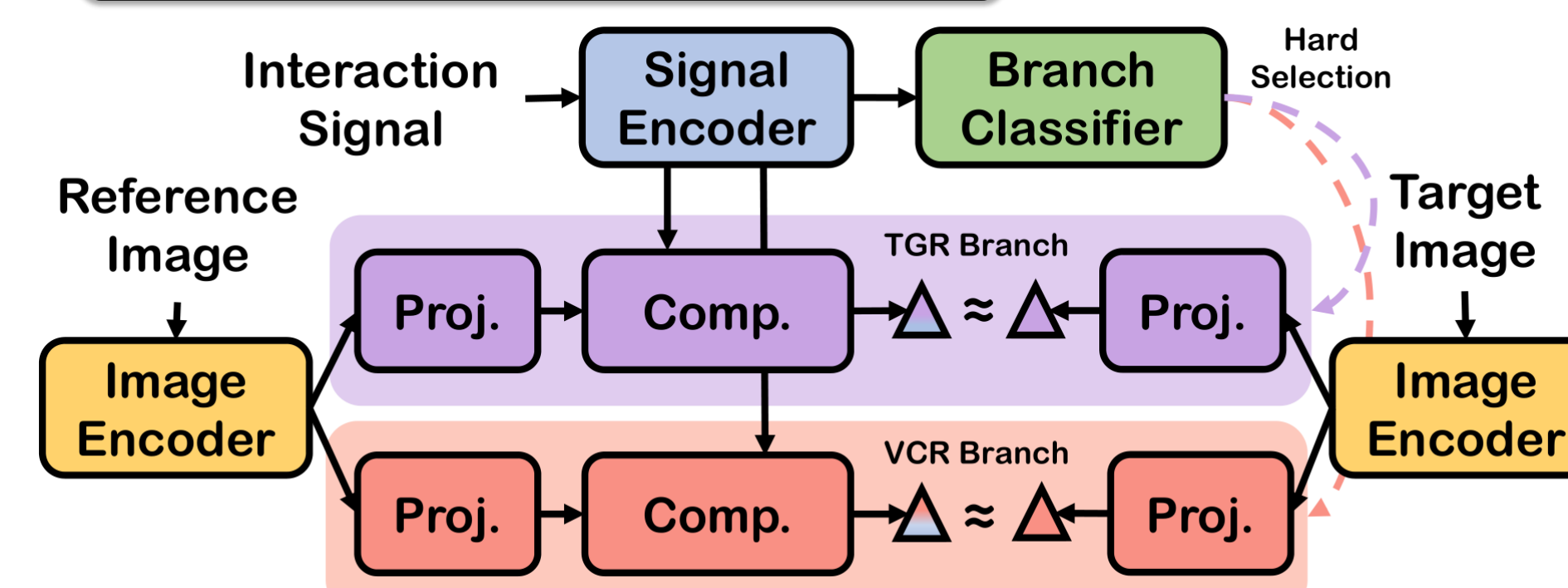
Benchmark Examples



Benchmark Statistics

Dataset	# Triplets	# Categories	Caption length
Shoes [3, 14]	10k	1	5.22 words
Fashion200K [16, 49]	172k	5	4.00 words
FashionIQ [51]	18k	3	5.36 words
Our TGR	381k	27	6.33 words
Dataset	# Outfits	# Categories	Interaction signal
Polyvore retrieval [30]	17k	16	One-hot labels
Our VCR	49k	27	Text

Model Architecture



Quantitative Results

Comp.	Metrics	TGR Results			VCR Results			Mean	
		R@10	R@50	mAP	R@10	R@50	mAP	R@K	mAP
CSA [30]	I	38.98	72.08	14.29	71.03	86.83	46.82	67.23	30.56
	U	36.90	70.57	13.37	70.46	86.88	46.47	66.20	29.92
TIRG [49]	I	46.27	77.57	19.78	69.30	85.88	46.15	69.76	32.97
	U	45.06	76.75	18.91	72.11	88.42	48.54	70.59	33.73
VAL [7]	I	43.27	75.10	18.06	62.99	81.97	40.40	65.83	29.23
	U	40.19	71.78	17.28	67.95	86.24	44.72	66.54	31.00
CoSMo [29]	I	40.24	72.15	17.31	64.10	83.18	41.36	64.92	29.34
	U	40.96	72.40	17.62	68.64	86.41	45.05	67.10	31.36
RTIC [44]	I	48.23	78.79	19.98	69.28	86.26	46.10	70.64	33.04
	U	46.75	77.80	19.26	74.18	89.56	50.78	72.07	35.02

Qualitative Results

