Mixture of Attribute-Aware Attention Experts for Fine-grained E-Commerce Composed Image Retrieval

Yufei Ma[1], Zihan Liang[1], Zhipeng Qian[1], Huangyu Dai, Lingtao Mao, Ben Chen[2], Chenyi Lei[2], Wenwu Ou

Kuaishou Technology [1] Equal contribution [2] Corresponding authors benchen4395@gmail.com

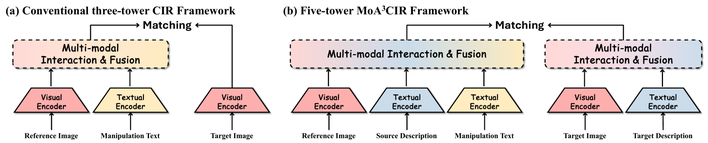
NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

Composed image retrieval combines a reference image with an edit request. E-commerce items also carry descriptions that reveal attributes such as material, brand and function.

What makes attribute edits difficult?

Real image pairs often change several attributes at once. This confounds supervision for a specific edit, while a single fused embedding can obscure fine-grained differences.



Five towers incorporate descriptions on both query and target sides.

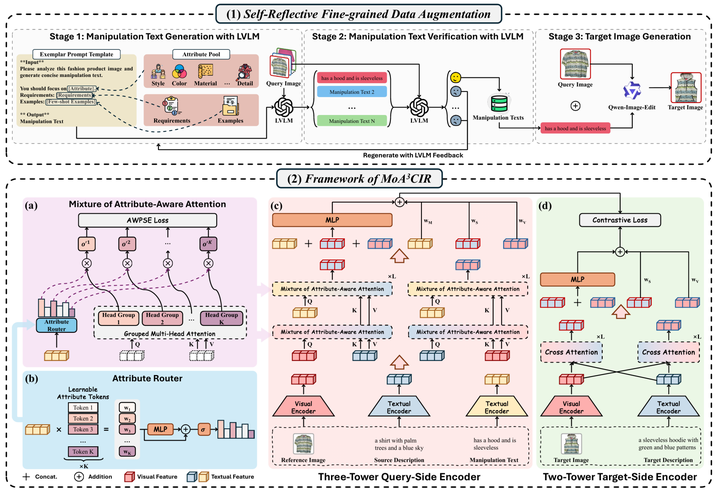
ECCIR benchmark

110K samples across 8 fashion and accessory categories, with authentic item descriptions from online logs.
92,160 training samples and 18,591 test samples.

Source Description	Reference Image	Manipulation Text	Target Image	Target Description
Plus-size rain boots for men, waterproof and non-slip.		*A **black low-top** version.*		New style men's rain boots, fishing casual water shoes, oil-proof, low-top short rain boots.
men's spring/autumn white denim jacket, korean style slim fit baseball jacket, cool brand workwear jacket.		*Do you have a **pink distressed denim** jacket with **no letters**?*		spring/autumn ripped denim jacket for teen boys, korean style lightweight loose denim coat.
yaloo women's lightweight short hooded slim fit down jacket with 90% white duck down, warm and lightweight coat.		*Do you have **gradient colors**?*		yaloo autumn/winter new lightweight white duck down jacket women gradient color short hooded fashion coat.
changxiang huili kids shoes white sports shoes autumn winter leather waterproof boy's rotating button white shoes soft sole.		***Mesh Netting** version please.*		changxiang huili boys' summer mesh white sneakers with rotating buttons, breathable upper, soft sole running shoes for kids.
Men's Spring and Autumn Washed and Distressed Jeans, American Style INS Retro Fashion Brand, Slightly Flared Pants.		*With **hole** and a **lighter color**.*		American high-street style, slim-flare jeans for men and women, vibe trendy brand, cool and edgy long pants.

Examples pair visual edits with real product descriptions.

MoA³CIR method



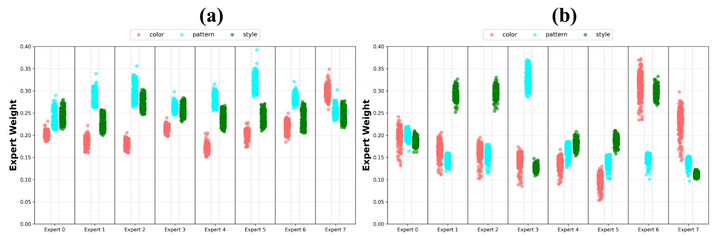
Attribute-aware attention experts

Grouped attention heads learn attribute subspaces. Learnable attribute tokens and sigmoid routing activate multiple experts according to the manipulation text.

Two-phase training

1. Specialize experts on controlled augmented edits with the asymmetric weighted point-set embedding (AWPSE) loss.

2. Freeze the MoA³ modules and optimize holistic query–target alignment on real data with InfoNCE.



Left: without Phase 1. Right: full training yields more distinct routing for color, pattern and style edits.

Retrieval results

Recall (%), higher is better. Selected published comparisons.

Method	ECCIR R@1	Shoes R@1
OFFSET	44.76	31.52
CIR-LVLM	52.72	31.40
MoA³CIR	**58.13**	**33.83**

+5.41 percentage points on ECCIR R@1

Compared with CIR-LVLM (52.72%).

Method	FashionIQ R@10	FashionIQ R@50
MAPNet	55.70	76.20
MoA³CIR	**57.51**	**77.34**

FashionIQ values average over Dress, Shirt and Toptee.

What drives the gains?

ECCIR configuration	R@1 (%)
Without MoA³ module	54.28
Without Phase 1	55.01
Without Phase 2	48.12
Full MoA³CIR	**58.13**

Scope

The largest benefits occur when item descriptions are available. ECCIR currently covers fashion and accessories from one platform. Augmentation adds a one-time model cost.