Kaleido-BERT: Vision-Language Pre-training on Fashion Domain

Mingchen Zhuge[1,*], Dehong Gao[1,*], Deng–Ping Fan[2,#], Linbo Jin[1], Ben Chen[1], Haoming Zhou[1], Minghui Qiu[1], Ling Shao[2]

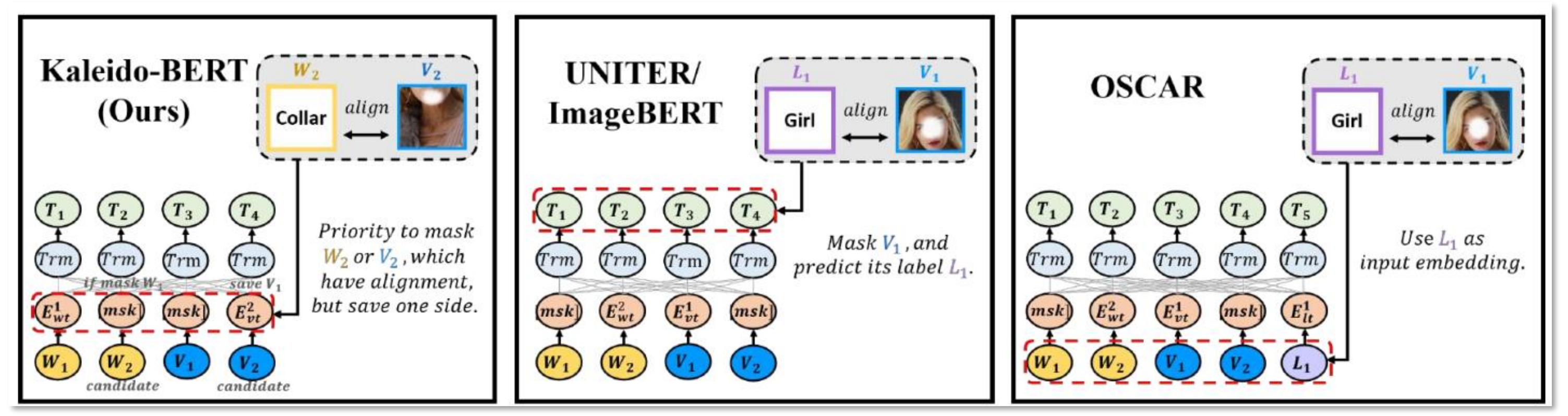
[1] Alibaba Group [2]Inception Institute of AI

CVPR 2021
VIRTUAL JUNE 19-25

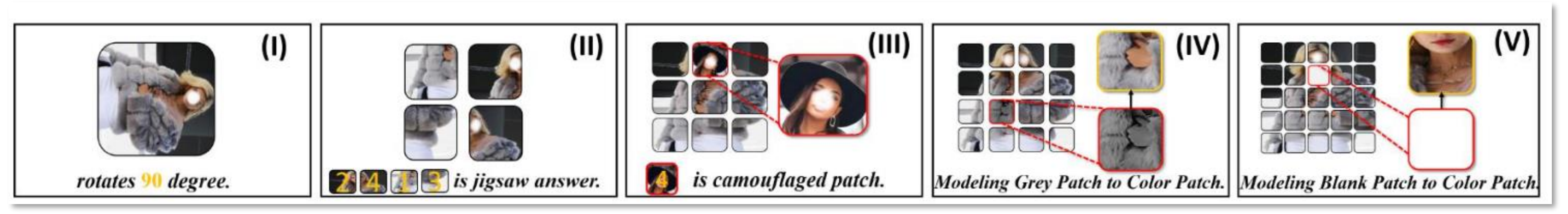
Alibaba Group 阿里巴巴集团
Inception Institute of Artificial Intelligence

1.Introduction

1. Different utilization of alignment information in VL pre-training architectures.



2. Aligned Kaleido Patch Modeling (AKPM).


(I) rotates 90 degree.
(II) 2 4 1 3 is jigsaw answer.
(III) 4 is camouflaged patch.
(IV) Modeling Grey Patch to Color Patch.
(V) Modeling Blank Patch to Color Patch.

Problems:

Existing VL pre-training models:

- Difficult to extend on specific domain.
- Without intelligent masking strategy.
- Use scale-fixed patches or RoIs as image inputs.
- Lack of image-level self-supervised pretext tasks.

Contributions:

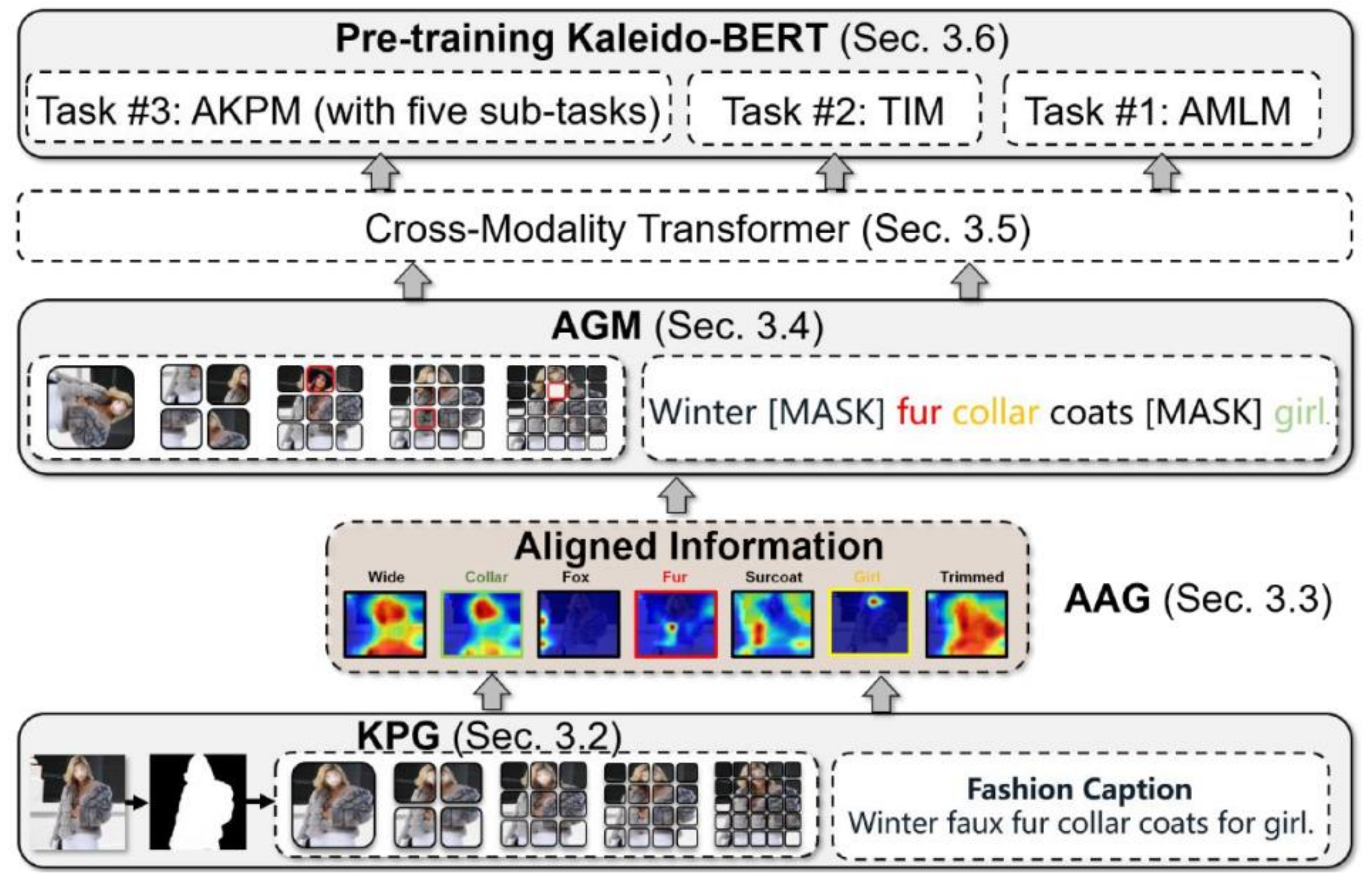
Based on the above problems, we

- Presented a strong **Kaleido-BERT** model.
- Design a useful **Alignment Guided Masking strategy**.
- Rethinking **5 self-supervised pretext tasks** in VL Pre-training process.

2.Vision-Languge Pre-training Model (Kaleido-BERT)

Kaleido-BERT, which consists:

- **KPG:** Kaleido Patch Generator
- **AGM:** Alignment Guided Masking
- **AAG:** Attention-based Alignment Generator
- **AKPM**: Aligned Kaleido Patch Modeling


Pre-training Kaleido-BERT (Sec. 3.6)
Task #3: AKPM (with five sub-tasks)
Task #2: TIM
Task #1: AMLM
Cross-Modality Transformer (Sec. 3.5)
AGM (Sec. 3.4)
Winter [MASK] fur collar coats [MASK] girl.
Aligned Information
Wide
Collar
Fox
Fur
Surcoat
Girl
Trimmed
AAG (Sec. 3.3)
KPG (Sec. 3.2)
Fashion Caption
Winter faux fur collar coats for girl.

3. Experiments (Tasks & Ablations)

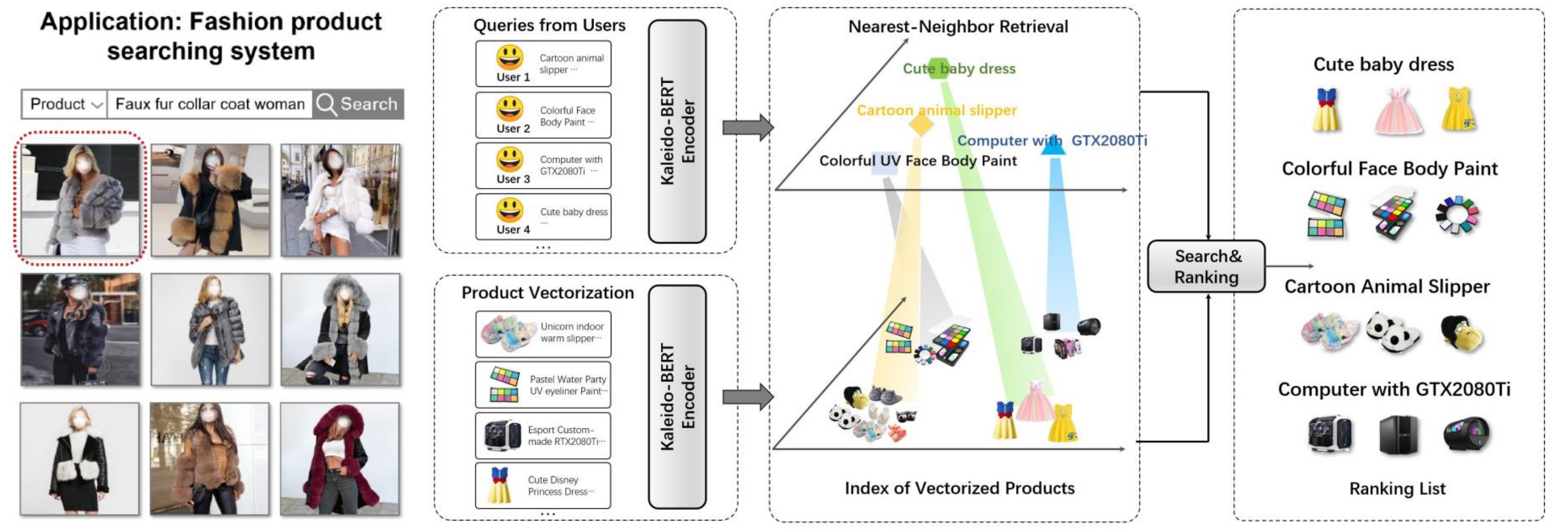
Tasks			ViLBERT [60]	VLBERT [45]	FashionBERT [21]	ImageBERT [55]	OSCAR [42]	Kaleido-BERT *Ours*
1.ITR	Rank@1	↑	20.97%	19.26%	23.96%	22.76%	23.39%	**27.99%** (+4.030%)
	Rank@5	↑	40.49%	39.90%	46.31%	41.89%	44.67%	**60.09%** (+13.78%)
	Rank@10	↑	48.21%	46.05%	52.12%	50.77%	52.55%	**68.37%** (+15.82%)
2.TIR	Rank@1	↑	21.12%	22.63%	26.75%	24.78%	25.10%	**33.88%** (+7.130%)
	Rank@5	↑	37.23%	36.48%	46.48%	45.20%	49.14%	**60.60%** (+11.46%)
	Rank@10	↑	50.11%	48.52%	55.74%	55.90%	56.68%	**68.59%** (+11.91%)
	Sum$\mathcal{R}$	↑	218.13	212.84	251.36	241.30	251.53	**319.52**

Tasks			FashionBERT [21]	ImageBERT [55]	OSCAR [42]	Kaleido-BERT *Ours*
3.CR	ACC	↑	91.25%	90.77%	91.79%	**95.07%** (+3.28%)
	macro-$\mathcal{F}$	↑	0.705	0.699	0.727	**0.714** (−0.013)
3.SUB	ACC	↑	85.27%	80.11%	84.23%	**88.07%** (+2.80%)
	macro-$\mathcal{F}$	↑	0.620	0.575	0.591	**0.636** (+0.016)
	Sum $\mathcal{CLS}$	↑	309.02	298.28	307.82	**318.14**
4.FC	Bleu-4	↑	3.30	-	4.50	**5.70** (+1.2)
	METEOR	↑	9.80	-	10.9	**12.8** (+1.9)
	ROUGE-L	↑	29.7	-	30.1	**32.9** (+2.8)
	CIDEr	↑	30.1	-	30.7	**32.6** (+1.9)
	Sum $\mathcal{CAP}$	↑	72.9	-	76.2	**84.0**

Metrics		KPG			AGM		AKPM						
		Scale-fixed	Kaleido.	Kaleido.+SOD	Random	AGM	B	B+I	B+I∼II	B+I∼III	B+I∼IV	B+I∼V	B+V
1. Rank@1	↑	24.71	26.73(+8.2%)	27.99(+13.3%)	26.55	27.99(+5.4%)	25.37	25.07(-1.2%)	26.03(+2.6%)	26.88(+6.0%)	26.20(+3.3%)	27.99(+10.3%)	24.62(-2.9%)
1. Rank@5	↑	50.05	54.55(+9.0%)	60.09(+20.1%)	55.13	60.09(+8.9%)	54.97	55.14(+0.3%)	56.31(+2.4%)	58.34(+6.1%)	59.13(+7.6%)	60.09(+9.3%)	53.78(-2.2%)
1. Rank@10	↑	58.93	65.44(+11.0%)	68.37(+16.0%)	64.92	68.37(+5.3%)	62.13	62.90(+1.2%)	63.37(+2.0%)	67.79(+9.1%)	67.99(+9.4%)	68.37(+10.0%)	60.88(-2.0)%
2. Rank@1	↑	30.17	32.19(+6.7%)	33.88(+12.0%)	32.14	33.88(+5.4%)	31.09	30.98(-0.4%)	32.22(+3.6%)	33.17(+6.7%)	33.80(+8.7%)	33.88(+9.0%)	30.77(-1.0%)
2. Rank@5	↑	52.29	58.40(+11.7%)	60.60(+15.9%)	56.99	60.60(+6.3%)	57.35	57.44(+0.2%)	58.73(+2.4%)	58.55(+2.1%)	60.57(+5.6%)	60.60(+5.7%)	55.95(-2.4%)
2. Rank@10	↑	60.82	66.49(+9.3%)	68.59(+12.8%)	63.77	68.59(+7.6%)	64.79	65.65(+1.3%)	64.16(-1.0%)	67.92(+4.8%)	68.41(+5.6%)	68.09(+5.1%)	61.70(-4.8%)
Sum $\mathcal{R}$	↑	**276.97**	**303.80(+9.7%)**	**319.52(+16.2%)**	**299.50**	**319.52(+6.7%)**	**295.70**	**297.18(+0.5%)**	**300.82(+1.7%)**	**312.65(+5.7%)**	**316.10(+6.9%)**	**319.02(+7.9%)**	**287.70(-2.7%)**
3. ACC	↑	93.44%	93.45%(+0.0%)	95.07%(+1.7%)	92.71%	95.07%(+2.5%)	90.94%	90.82%(-0.1%)	91.40%(+0.5%)	93.91%(+3.3%)	94.05%(+3.4%)	95.07%(+4.5%)	88.87(-2.3%)
3. macro-$\mathcal{F}$	↑	0.701	0.705(+0.6%)	0.714(+1.9%)	0.711	0.714(+0.4%)	0.690	0.692(+0.3%)	0.721(+4.5%)	0.713(+3.3%)	0.710(+2.9%)	0.714(+3.5%)	0.701(+1.4%)
4. ACC	↑	86.89%	87.61%(+0.8%)	88.07%(+1.4%)	87.20%	88.07(+1.0%)	81.66%	81.25%(-0.5%)	84.44%(+3.4%)	86.49%(+5.9%)	88.53%(+8.4%)	88.07%(+7.9%)	81.64(+0.0%)
4. macro-$\mathcal{F}$	↑	0.630	0.634(+0.6%)	0.636(+1.0%)	0.633	0.636(+0.5%)	0.558	0.575(+3.0%)	0.596(+6.8%)	0.636(+14.0%)	0.633(+13.4%)	0.636(+14.0%)	0.596(+8.4%)
Sum $\mathcal{CLS}$	↑	**313.43**	**314.96(+0.5%)**	**318.14(+1.5%)**	**314.31**	**318.14(+1.2%)**	**297.40**	**298.77(+0.4%)**	**307.54(+3.4%)**	**315.30(+6.0%)**	**316.88(+6.5%)**	**318.14(+7.0%)**	**300.21(+0.9%)**
5. Bleu-4	↑	4.9	5.2(+6.1%)	5.7(+16.3%)	5.3	5.7(+7.5%)	4.9	5.2(+6.1%)	5.2(+6.1%)	5.1(+4.1%)	5.6(+14.3%)	5.7(+16.3%)	5.3(+8.2%)
5. METEOR	↑	11.0	11.7(+6.4%)	12.8(+16.4%)	11.3	12.8(+13.3%)	11.6	11.6(+0.0%)	11.8(+1.7%)	12.6(+8.6%)	12.8(+10.3%)	12.8(+10.3%)	11.4(-1.7%)
5. ROUGE-L	↑	29.8	31.5(+5.7%)	32.9(+10.4%)	30.3	32.9(+8.6%)	30.4	30.7(+1.0%)	30.8(+1.3%)	31.9(+4.9%)	32.7(+7.6%)	32.9(+8.2%)	30.6(+0.7%)
5. CIDEr	↑	30.9	31.3(+1.3%)	32.6(+5.5%)	31.7	32.6(+2.8%)	31.0	31.5(+1.6%)	31.4(+1.3%)	32.0(+3.2%)	32.3(+4.2%)	32.6(+5.2%)	31.3(+1.0%)
Sum $\mathcal{CAP}$	↑	**76.6**	**79.7(+4.0%)**	**84.0(+9.7%)**	**78.6**	**84.0(+6.9%)**	**77.9**	**79.0(+1.4%)**	**79.2(+1.7%)**	**81.6(+4.7%)**	**83.4(+7.1%)**	**84.0(+7.8%)**	**78.6(+0.9%)**

4.Results

5.Application (Searching System)


Application: Fashion product searching system
Product
Faux fur collar coat woman
Search
Queries from Users
User 1 Cartoon animal slipper ···
User 2 Colorful Face Body Paint ···
User 3 Computer with GTX2080Ti ···
User 4 Cute baby dress ···
Kaleido-BERT Encoder
Product Vectorization
Unicorn indoor warm slipper···
Pastel Water Party UV eyeliner Paint···
Esport Custom-made RTX2080Ti···
Cute Disney Princess Dress···
Kaleido-BERT Encoder
Nearest-Neighbor Retrieval
Cute baby dress
Cartoon animal slipper
Computer with GTX2080Ti
Colorful UV Face Body Paint
Index of Vectorized Products
Search& Ranking
Cute baby dress
Colorful Face Body Paint
Cartoon Animal Slipper
Computer with GTX2080Ti
Ranking List

Cross-modality Semantic Searching at Alibaba.com.

6.More Information

https://dpfan.net/Kaleido-BERT
https://github.com/mczhuge/Kaleido-BERT
dengpfan@gmail.com
mczhuge@gmail.com

Wechat