SHANDONG UNIVERSITY 1901

SIGIR 2022
Madrid

達摩院
ALIBABA DAMO ACADEMY

V2P: Vision-to-Prompt based Multi-Modal Product Summary Generation

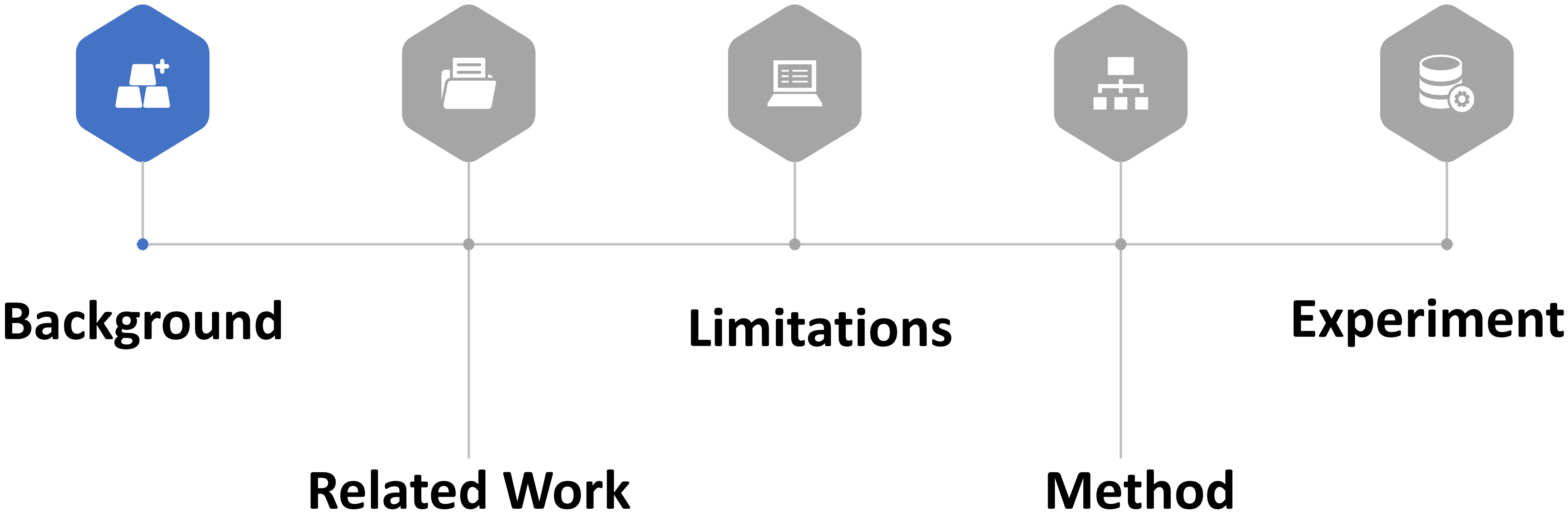
Xuemeng Song[†], Liqiang Jing[†], Dengtian Lin[†],
Zhongzhou Zhao[§], Haiqing Chen[§], Liqiang Nie[†]

[†]Shandong University, Shandong, China,
[§]Alibaba Group, Hangzhou, China

Presenter: Liqiang Jing

jingliqiang6@gmail.com

Outline



Background

Related Work

Limitations

Method

Experiment

Background

Online shopping platform always shows various information about the product.

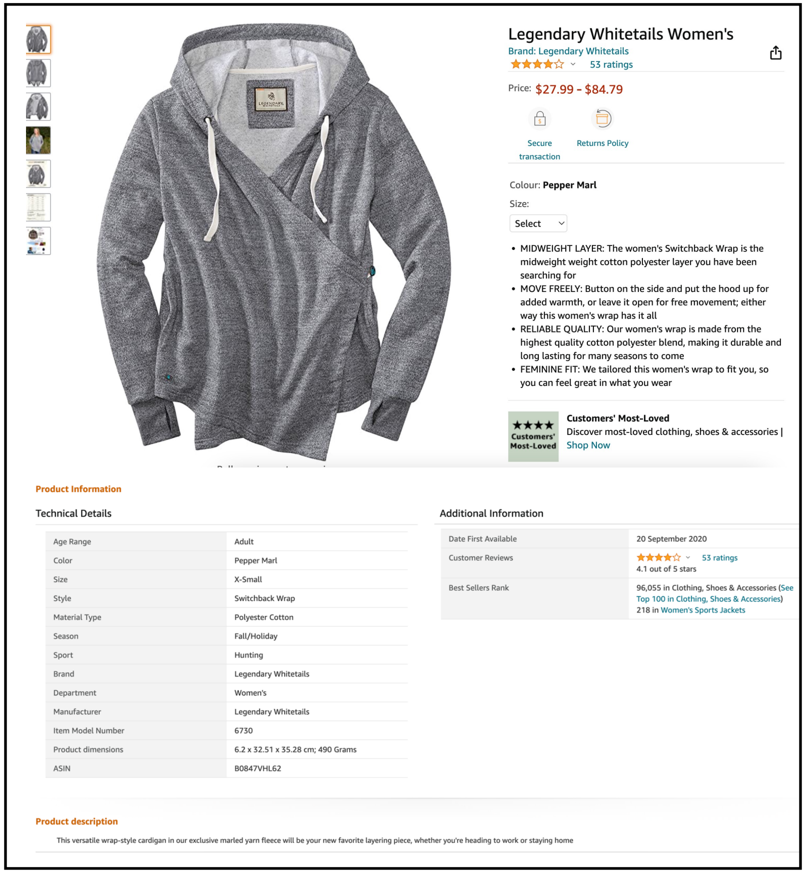
OBERLO

How Many People Shop Online?

An estimated 2.14 BILLION people worldwide purchase goods online.

(Statista, 2017)

https://www.oberlo.com/blog/online-shopping-statistics



Legendary Whitetails Women's

Brand: Legendary Whitetails

★★★★☆ 53 ratings

Price: $27.99 - $84.79

Colour: **Pepper Marl**

Size:

Select

- MIDWEIGHT LAYER: The women's Switchback Wrap is the midweight weight cotton polyester layer you have been searching for
- MOVE FREELY: Button on the side and put the hood up for added warmth, or leave it open for free movement; either way this women's wrap has it all
- RELIABLE QUALITY: Our women's wrap is made from the highest quality cotton polyester blend, making it durable and long lasting for many seasons to come
- FEMININE FIT: We tailored this women's wrap to fit you, so you can feel great in what you wear

Customers' Most-Loved
Discover most-loved clothing, shoes & accessories | Shop Now

Product Information

Technical Details

Age Range	Adult
Color	Pepper Marl
Size	X-Small
Style	Switchback Wrap
Material Type	Polyester Cotton
Season	Fall/Holiday
Sport	Hunting
Brand	Legendary Whitetails
Department	Women's
Manufacturer	Legendary Whitetails
Item Model Number	6730
Product dimensions	6.2 x 32.51 x 35.28 cm; 490 Grams
ASIN	B0847VHL62

Additional Information

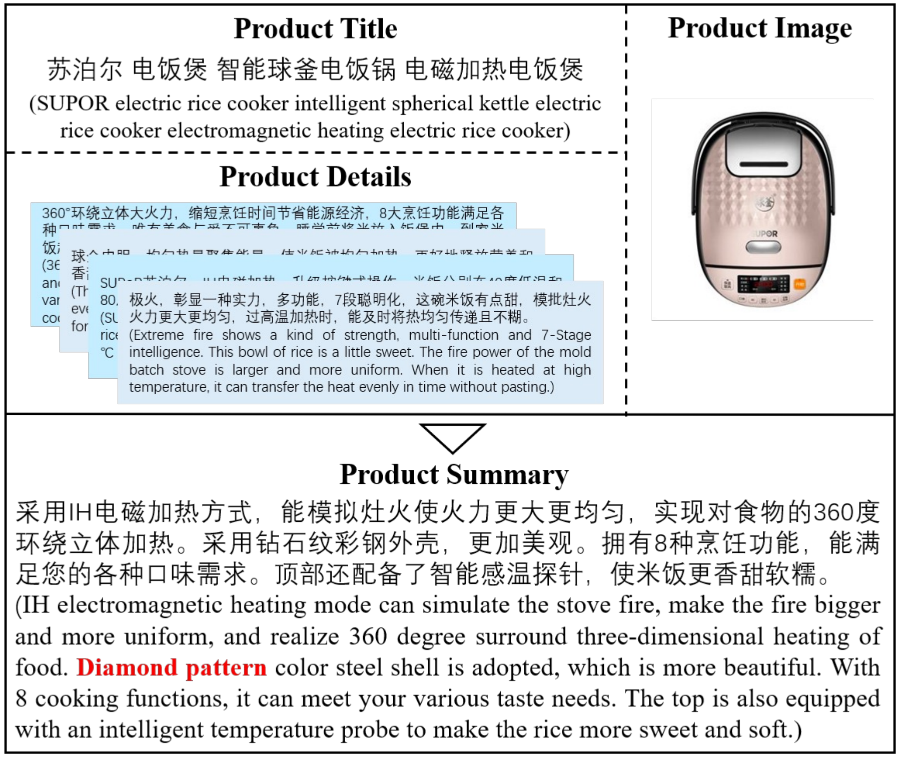
Date First Available	20 September 2020
Customer Reviews	★★★★☆ 53 ratings 4.1 out of 5 stars
Best Sellers Rank	96,055 in Clothing, Shoes & Accessories (See Top 100 in Clothing, Shoes & Accessories) 218 in Women's Sports Jackets

Product description

This versatile wrap-style cardigan in our exclusive marled yarn fleece will be your new favorite layering piece, whether you're heading to work or staying home

Background

- Multi-Modal Product Summarization Generation (MPSG)



Product Title

苏泊尔 电饭煲 智能球釜电饭锅 电磁加热电饭煲
(SUPOR electric rice cooker intelligent spherical kettle electric rice cooker electromagnetic heating electric rice cooker)

Product Details

360°环绕立体大火力，缩短烹饪时间节省能源经济，8大烹饪功能满足各

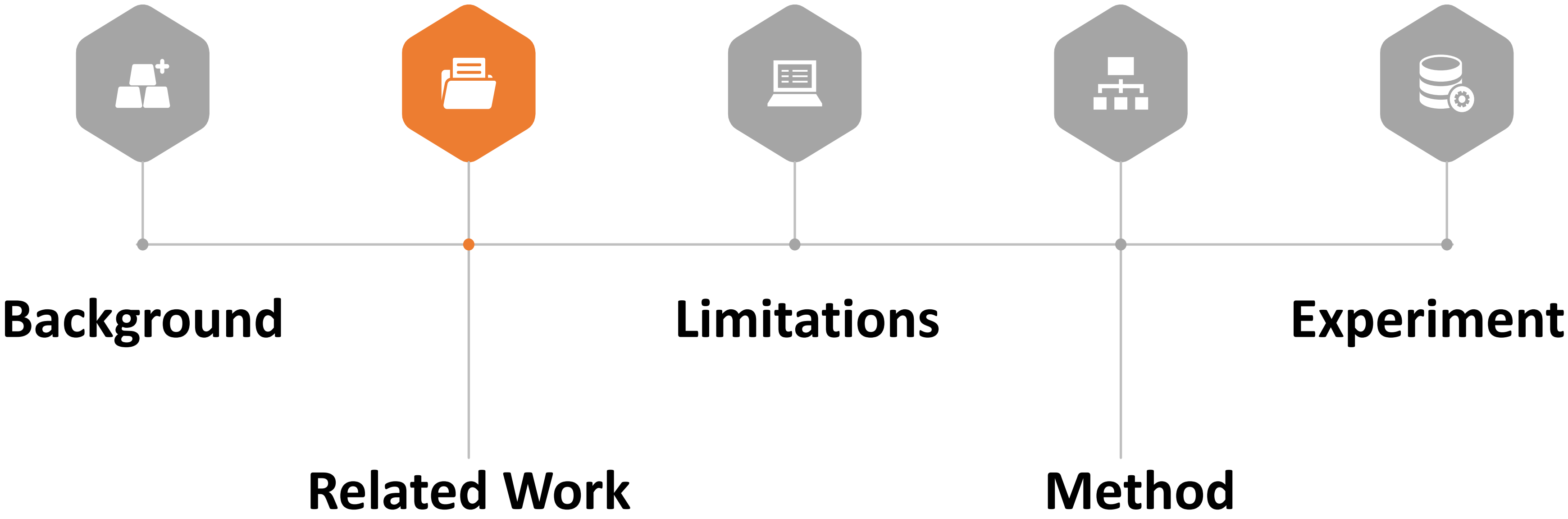
极火，彰显一种实力，多功能，7段聪明化，这碗米饭有点甜，模批灶火火力更大更均匀，过高温加热时，能及时将热均匀传递且不糊。
(Extreme fire shows a kind of strength, multi-function and 7-Stage intelligence. This bowl of rice is a little sweet. The fire power of the mold batch stove is larger and more uniform. When it is heated at high temperature, it can transfer the heat evenly in time without pasting.)

Product Image

Product Summary

采用IH电磁加热方式，能模拟灶火使火力更大更均匀，实现对食物的360度环绕立体加热。采用钻石纹彩钢外壳，更加美观。拥有8种烹饪功能，能满足您的各种口味需求。顶部还配备了智能感温探针，使米饭更香甜软糯。
(IH electromagnetic heating mode can simulate the stove fire, make the fire bigger and more uniform, and realize 360 degree surround three-dimensional heating of food. **Diamond pattern** color steel shell is adopted, which is more beautiful. With 8 cooking functions, it can meet your various taste needs. The top is also equipped with an intelligent temperature probe to make the rice more sweet and soft.)

Outline



Background

Related Work

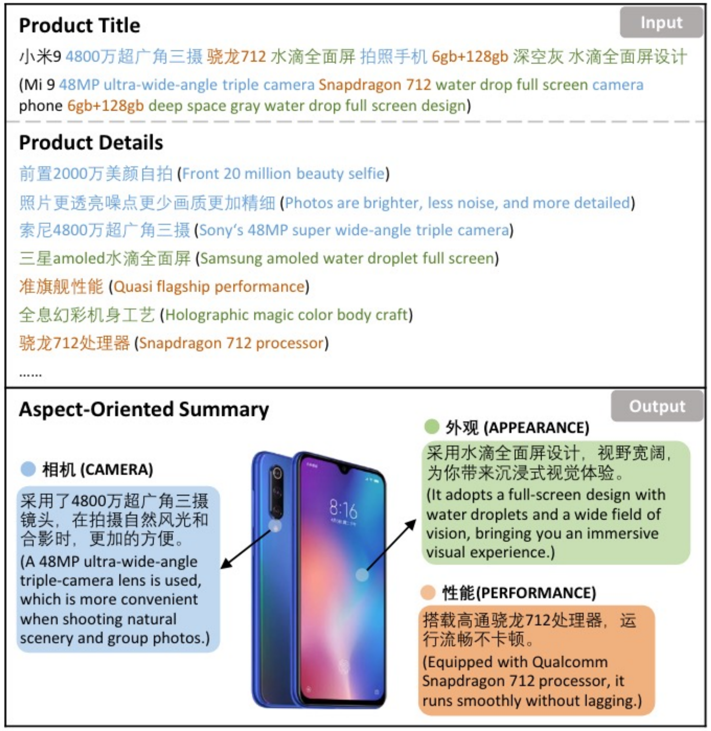
Limitations

Method

Experiment

Related Work

- ➢ Traditional Product Summarization



Product Title Input

小米9 4800万超广角三摄 骁龙712 水滴全面屏 拍照手机 6gb+128gb 深空灰 水滴全面屏设计

(Mi 9 48MP ultra-wide-angle triple camera Snapdragon 712 water drop full screen camera phone 6gb+128gb deep space gray water drop full screen design)

Product Details

前置2000万美颜自拍 (Front 20 million beauty selfie)

照片更透亮噪点更少画质更加精细 (Photos are brighter, less noise, and more detailed)

索尼4800万超广角三摄 (Sony's 48MP super wide-angle triple camera)

三星amoled水滴全面屏 (Samsung amoled water droplet full screen)

准旗舰性能 (Quasi flagship performance)

全息幻彩机身工艺 (Holographic magic color body craft)

骁龙712处理器 (Snapdragon 712 processor)

……

Aspect-Oriented Summary Output

● 相机 (CAMERA)

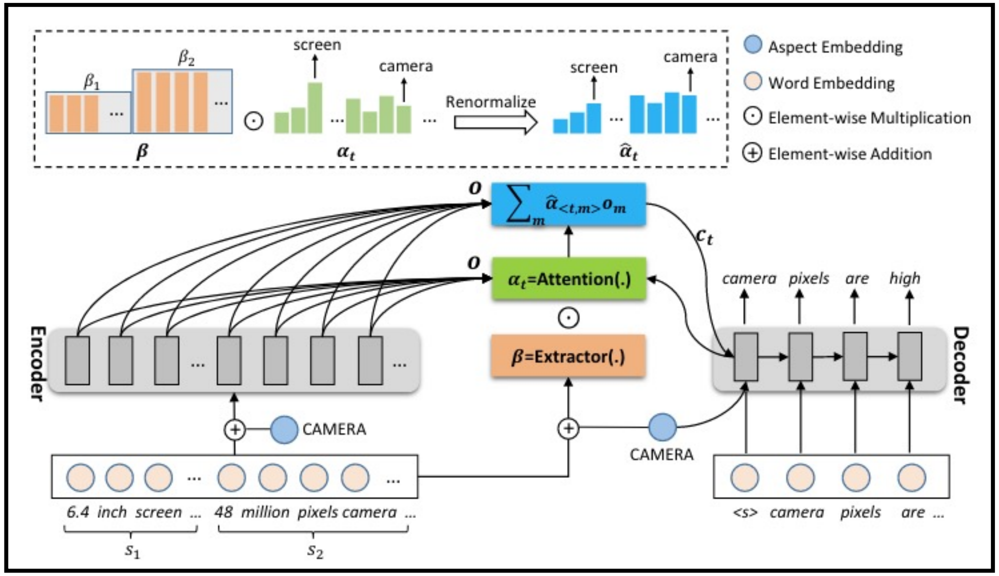
采用了4800万超广角三摄镜头，在拍摄自然风光和合影时，更加的方便。
(A 48MP ultra-wide-angle triple-camera lens is used, which is more convenient when shooting natural scenery and group photos.)

● 外观 (APPEARANCE)

采用水滴全面屏设计，视野宽阔，为你带来沉浸式视觉体验。
(It adopts a full-screen design with water droplets and a wide field of vision, bringing you an immersive visual experience.)

● 性能(PERFORMANCE)

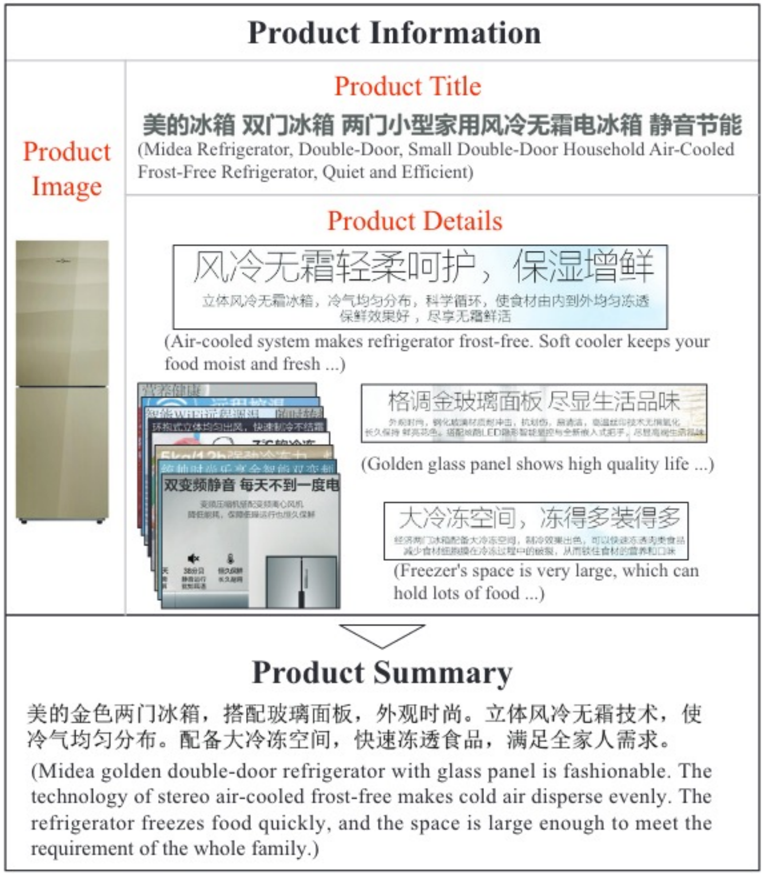
搭载高通骁龙712处理器，运行流畅不卡顿。
(Equipped with Qualcomm Snapdragon 712 processor, it runs smoothly without lagging.)



Jiahui Liang et al. CUSTOM: Aspect-Oriented Product Summarization for E-Commerce. NLPCC. 2021.

Related Work

- ➢ Multi-Modal Product Summarization



Product Information

Product Image

Product Title

美的冰箱 双门冰箱 两门小型家用风冷无霜电冰箱 静音节能

(Midea Refrigerator, Double-Door, Small Double-Door Household Air-Cooled Frost-Free Refrigerator, Quiet and Efficient)

Product Details

风冷无霜轻柔呵护，保湿增鲜

立体风冷无霜冰箱，冷气均匀分布，科学循环，使食材由内到外均匀冻透 保鲜效果好，尽享无霜鲜活

(Air-cooled system makes refrigerator frost-free. Soft cooler keeps your food moist and fresh ...)

双变频静音 每天不到一度电

格调金玻璃面板 尽显生活品味

(Golden glass panel shows high quality life ...)

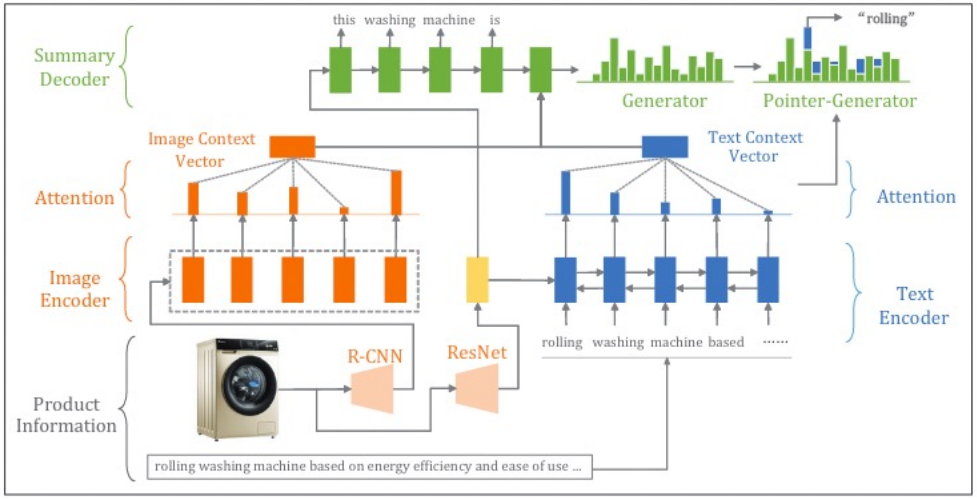
大冷冻空间，冻得多装得多

(Freezer's space is very large, which can hold lots of food ...)

Product Summary

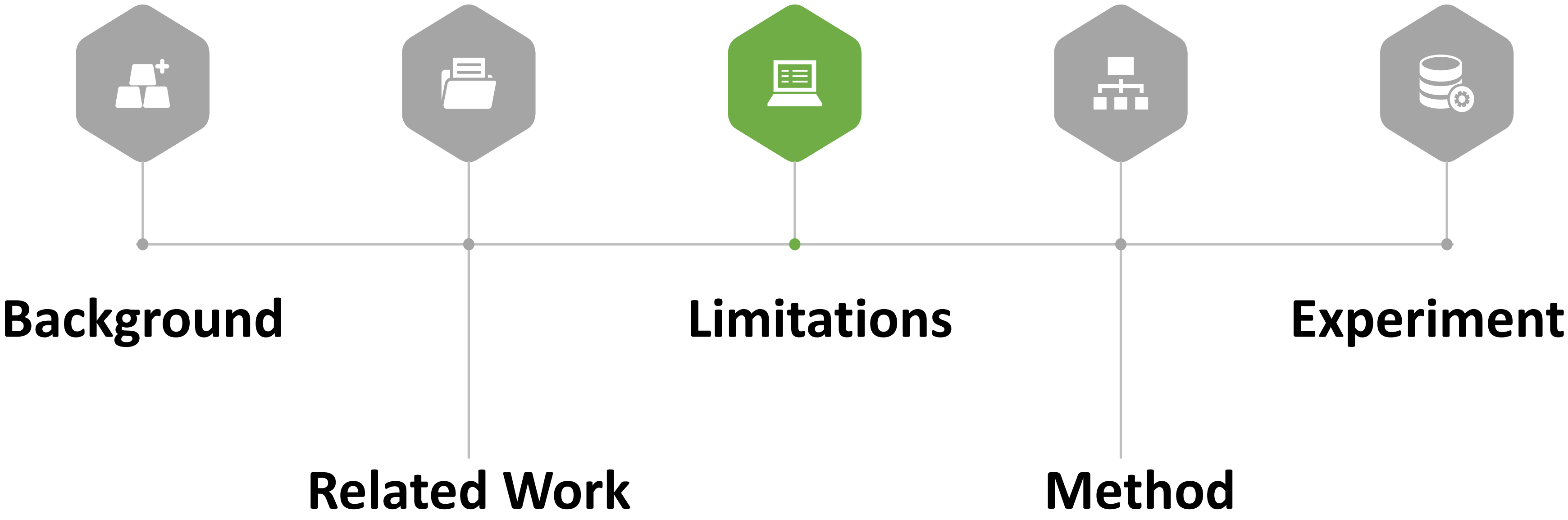
美的金色两门冰箱，搭配玻璃面板，外观时尚。立体风冷无霜技术，使冷气均匀分布。配备大冷冻空间，快速冻透食品，满足全家人需求。

(Midea golden double-door refrigerator with glass panel is fashionable. The technology of stereo air-cooled frost-free makes cold air disperse evenly. The refrigerator freezes food quickly, and the space is large enough to meet the requirement of the whole family.)



Haoran Li et al. Aspect-Aware Multimodal Summarization for Chinese E-Commerce Products. AAAI. 2020.

Outline



Background

Related Work

Limitations

Method

Experiment

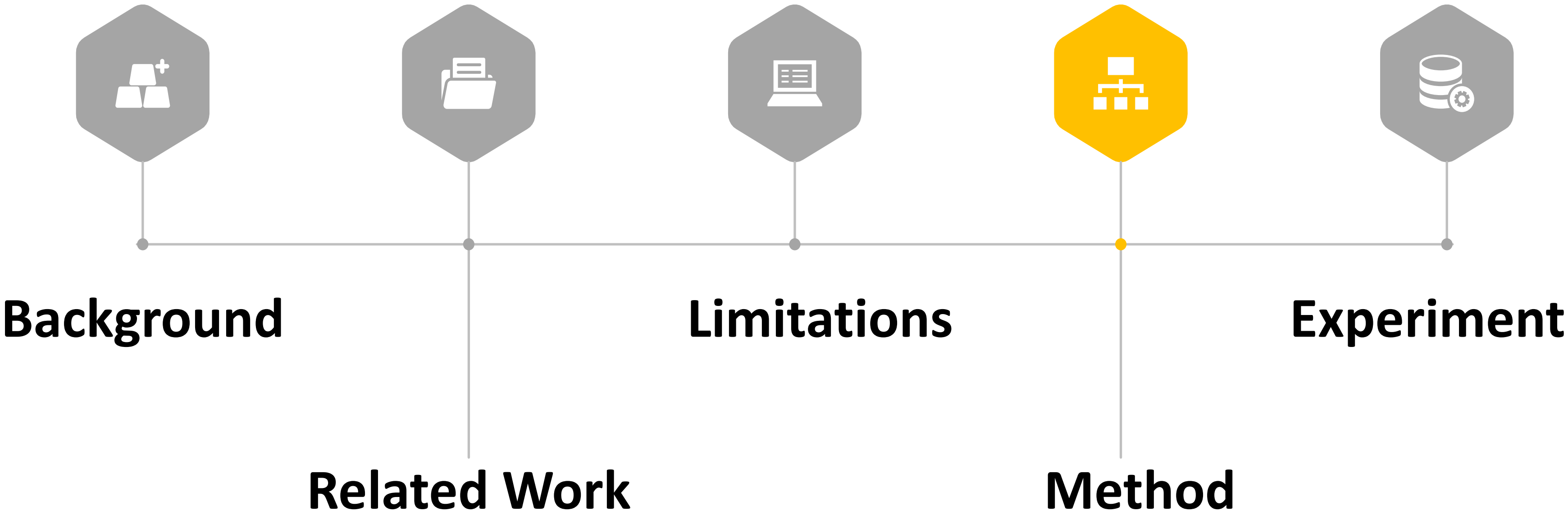
Limitations

- L1: Overlook the benefit of pre-training.

- L2: Lack the representation-level supervision.

- L3: Ignore the diversity of the seller-generated data.

Outline



Background

Related Work

Limitations

Method

Experiment

Method

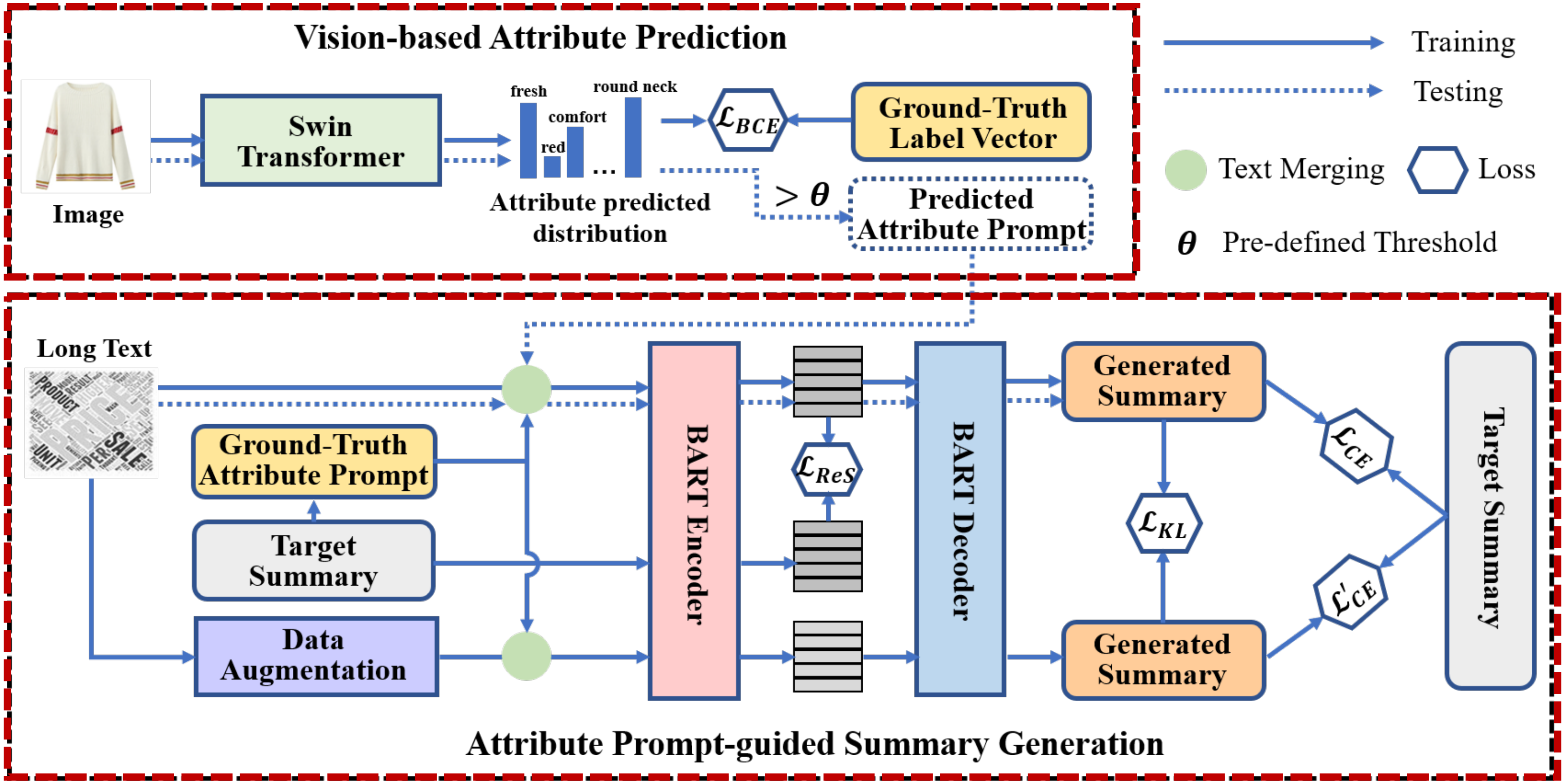
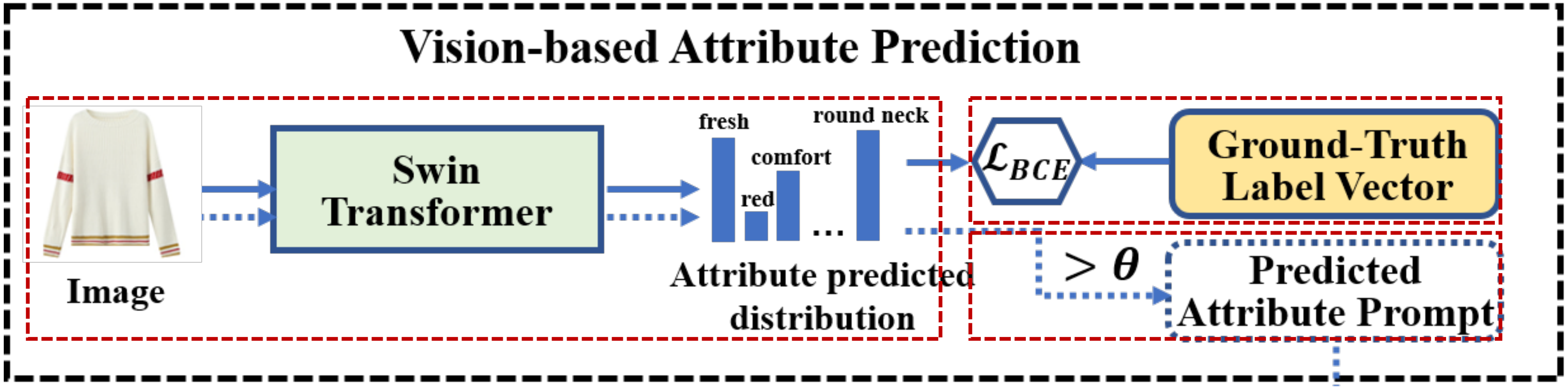


Fig.1: The proposed V2P scheme.

Method


Vision-based Attribute Prediction
Image
Swin Transformer
fresh
comfort
red
round neck
...
Attribute predicted distribution
$\mathcal{L}_{BCE}$
Ground-Truth Label Vector
$> \theta$
Predicted Attribute Prompt

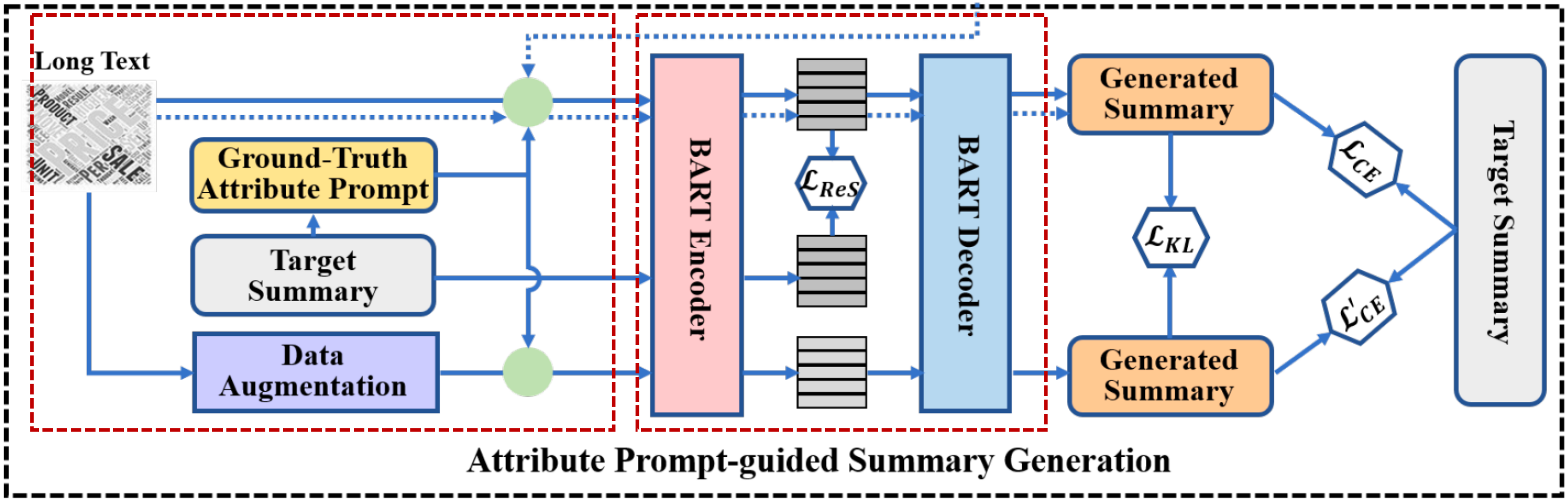
- Attribute Prediction

$$\hat{\mathbf{s}} = Swin(V),$$

$$\widehat{\mathcal{A}} = \{a^k | \hat{s}^k \geq \theta, k = 1, \cdots, S\},$$

$$L_{BCE} = \min_{\Theta_a} -[\mathbf{s} \ln(\sigma(\hat{\mathbf{s}})) + (\mathbf{1} - \mathbf{s}) \ln(\mathbf{1} - \sigma(\hat{\mathbf{s}}))],$$

Method



Attribute Prompt-guided Summary Generation

Text and Attribute Prompt Embedding

$$\widehat{X} = [X, a^{k_1}, a^{k_2}, \cdots, a^{k_Q}]$$

$$\mathbf{e}^j = \mathbf{W}^T \mathbf{g}^j, j = 1, \cdots, U,$$

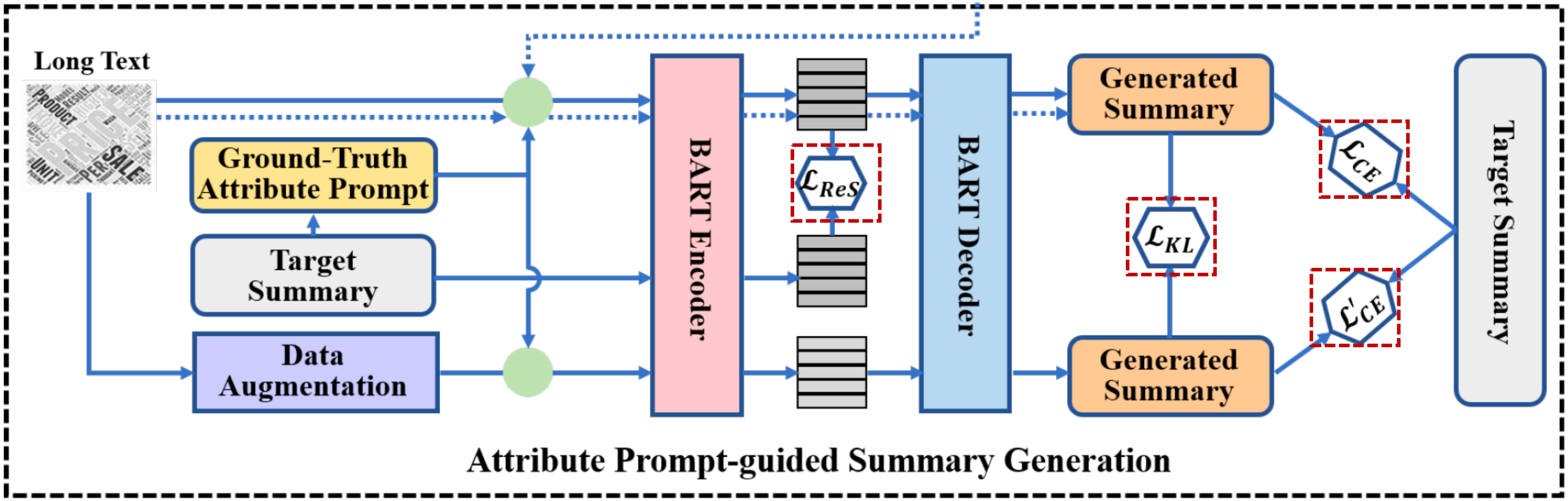
$$\mathbf{E} = [\mathbf{e}^1; \mathbf{e}^2; \cdots; \mathbf{e}^U] + \mathbf{E}_{pos},$$

BART-based Summary Generation

$$\mathrm{Z} = \mathcal{E}(\mathrm{E}),$$

$$\hat{\mathrm{p}}_\mathrm{j} = \mathcal{D}(\mathrm{Z}, \hat{y}_1, \hat{y}_2, \cdots, \hat{y}_{j-1}),$$

Method



Attribute Prompt-guided Summary Generation

Output-level and Representation-level Supervision

$$\mathcal{L}_{CE} = -\frac{1}{L}\sum_{j=1}^{L} log(\hat{\mathrm{p}}_j[t*]),$$

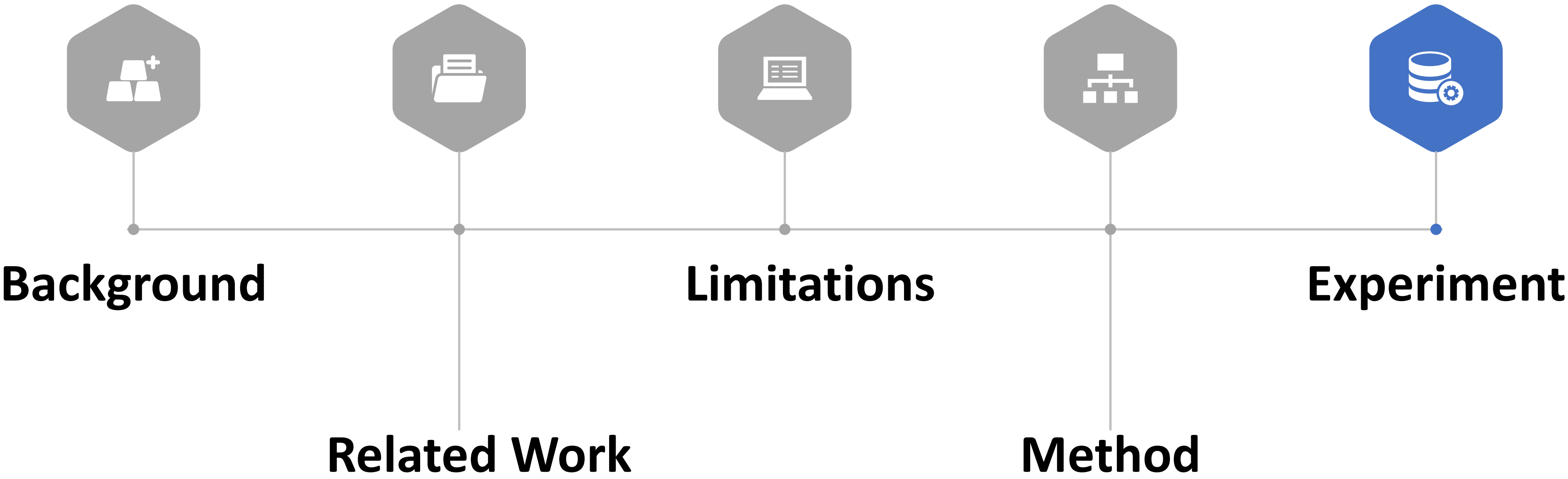
$$\mathcal{L}_{ReS} = \mathbb{E}_{p(z_{1:B})p(\tilde{z}|z_1)}\Big[-log\frac{e^{f(z^y,z_1)}}{\sum_{i=1}^{B} e^{f(z^y,z_i)}}\Big],$$

Data Augmentation-based Robustness Regularization

$$\mathcal{L}'_{CE} = -\frac{1}{L}\sum_{i=1}^{L} log(\hat{\mathrm{p}}'_j[t*]),$$

$$\mathcal{L}_{KL} = \sum_{j=1}^{L} D_{KL}(\hat{\mathrm{p}}_j\|\hat{\mathrm{p}}'_j) = \sum_{j=1}^{L}\sum_{k=1}^{L}\hat{p}_{jk}log(\frac{\hat{p}_{jk}}{\hat{p}'_{jk}}).$$

Outline



Background

Related Work

Limitations

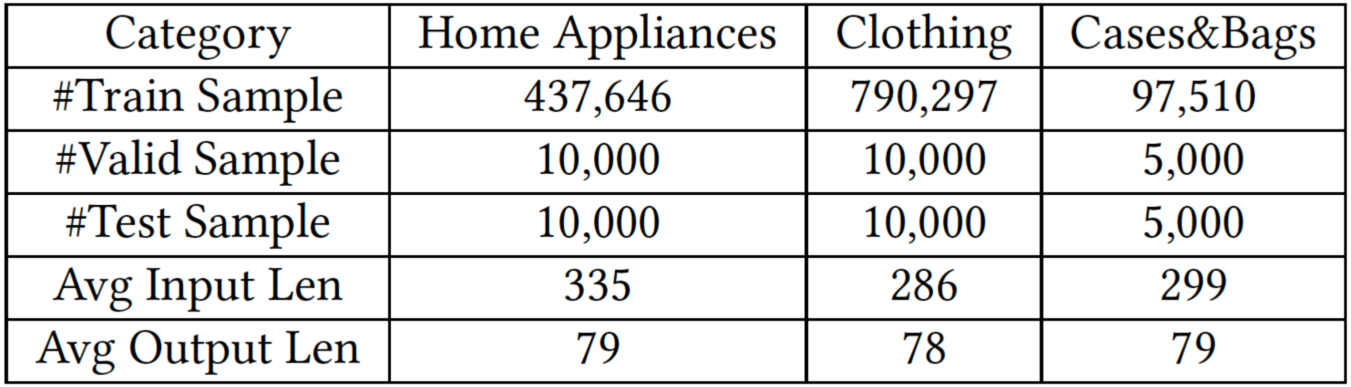
Method

Experiment

Experiment

Dataset

Table 1: Dataset statistics of CEPSUM.



Category	Home Appliances	Clothing	Cases&Bags
#Train Sample	437,646	790,297	97,510
#Valid Sample	10,000	10,000	5,000
#Test Sample	10,000	10,000	5,000
Avg Input Len	335	286	299
Avg Output Len	79	78	79

Evaluation Metric

Rouge-1, Rouge-2, Rouge-L.

Experiment

➢ On Model Comparison

Table 2: Performance comparison among different methods.

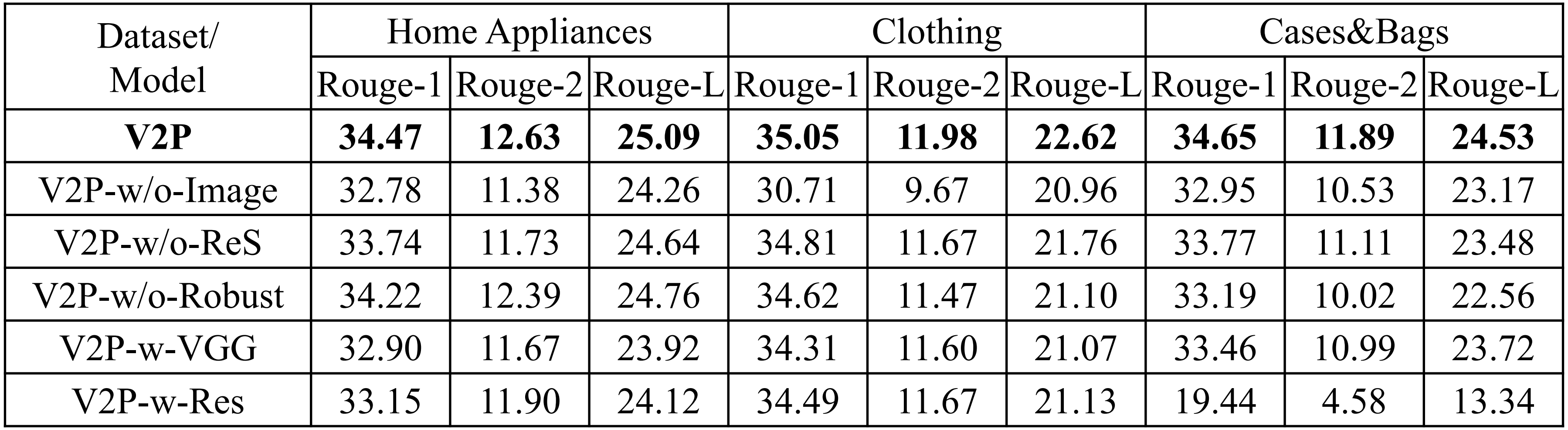
Dataset/ Model	Home Appliances			Clothing			Cases&Bags		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
Lead	21.97	9.54	12.79	19.83	8.39	13.56	21.49	9.37	14.19
LexBank	24.06	10.01	18.19	26.87	9.01	17.76	27.09	9.87	18.03
Seq2Seq	21.57	7.18	17.61	23.05	6.84	16.82	23.18	6.94	17.29
MASS	28.19	8.02	18.73	26.73	8.03	17.72	27.19	9.03	18.17
PG	31.31	10.93	21.11	29.11	9.24	19.92	31.11	10.27	21.79
MMPG	32.88	11.88	21.96	30.73	10.29	21.25	32.69	11.78	22.27
VG-Bart-1	32.71	11.46	22.87	31.36	9.94	21.34	32.73	10.26	22.44
VG-Bart-2	32.73	11.74	23.61	31.63	10.08	21.64	33.30	11.31	23.13
V2P	**34.47***	**12.63***	**25.09***	**35.05***	**11.98***	**22.62***	**34.65***	**11.89***	**24.53***
Improvement . ↑	4.84%	6.32%	6.27%	10.81%	16.42%	5.06%	4.05%	0.93%	6.05%

V2P consistently surpasses all the baselines, exhibiting the effectiveness of the proposed scheme.

Experiment

- On Ablation Study

Table 3: Performance comparison between V2P and its derivatives.

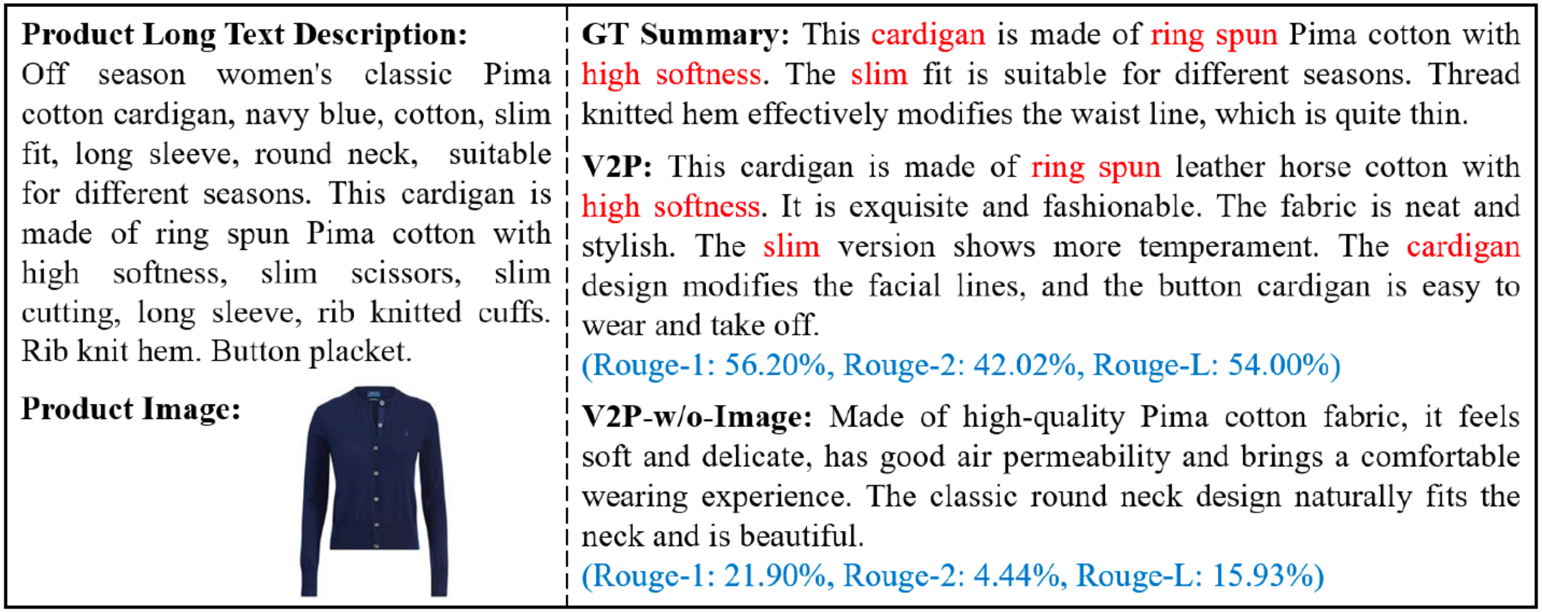


Dataset/ Model	Home Appliances			Clothing			Cases&Bags		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
V2P	**34.47**	**12.63**	**25.09**	**35.05**	**11.98**	**22.62**	**34.65**	**11.89**	**24.53**
V2P-w/o-Image	32.78	11.38	24.26	30.71	9.67	20.96	32.95	10.53	23.17
V2P-w/o-ReS	33.74	11.73	24.64	34.81	11.67	21.76	33.77	11.11	23.48
V2P-w/o-Robust	34.22	12.39	24.76	34.62	11.47	21.10	33.19	10.02	22.56
V2P-w-VGG	32.90	11.67	23.92	34.31	11.60	21.07	33.46	10.99	23.72
V2P-w-Res	33.15	11.90	24.12	34.49	11.67	21.13	19.44	4.58	13.34

V2P obtains the best performance, which verifies these components are significant in our model.

Experiment

➢ On Case Study



Product Long Text Description: Off season women's classic Pima cotton cardigan, navy blue, cotton, slim fit, long sleeve, round neck, suitable for different seasons. This cardigan is made of ring spun Pima cotton with high softness, slim scissors, slim cutting, long sleeve, rib knitted cuffs. Rib knit hem. Button placket.

Product Image:

GT Summary: This cardigan is made of ring spun Pima cotton with high softness. The slim fit is suitable for different seasons. Thread knitted hem effectively modifies the waist line, which is quite thin.

V2P: This cardigan is made of ring spun leather horse cotton with high softness. It is exquisite and fashionable. The fabric is neat and stylish. The slim version shows more temperament. The cardigan design modifies the facial lines, and the button cardigan is easy to wear and take off.
(Rouge-1: 56.20%, Rouge-2: 42.02%, Rouge-L: 54.00%)

V2P-w/o-Image: Made of high-quality Pima cotton fabric, it feels soft and delicate, has good air permeability and brings a comfortable wearing experience. The classic round neck design naturally fits the neck and is beautiful.
(Rouge-1: 21.90%, Rouge-2: 4.44%, Rouge-L: 15.93%)

Figure 2: Comparison between the summaries generated by our model and its variant V2P-w/o-image for a clothing product.

Conclusion

- We design a **vision-to-prompt** based multi-modal product summary generation scheme, where the **heterogeneous multi-modal data** are unified in the same space by converting the vision modality into **semantic attribute prompts**.

- We are the first to perform both **output-level** and **representation-level** supervision for MPSG simultaneously and introduce the **data augmentation-based** robustness regularization.

- We are the first to adapt the **generative pre-trained language model** to solve the MPSG task and achieve superior performance on a real-world dataset.

Thanks for your listening.

Codes are available!