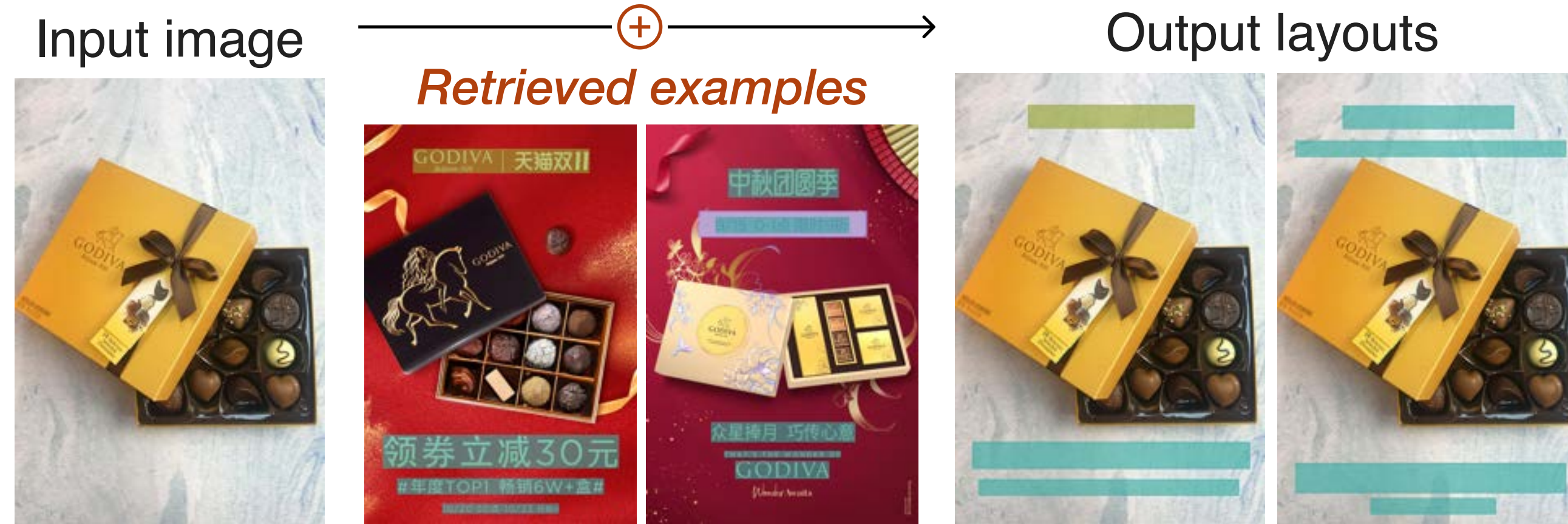


Concept

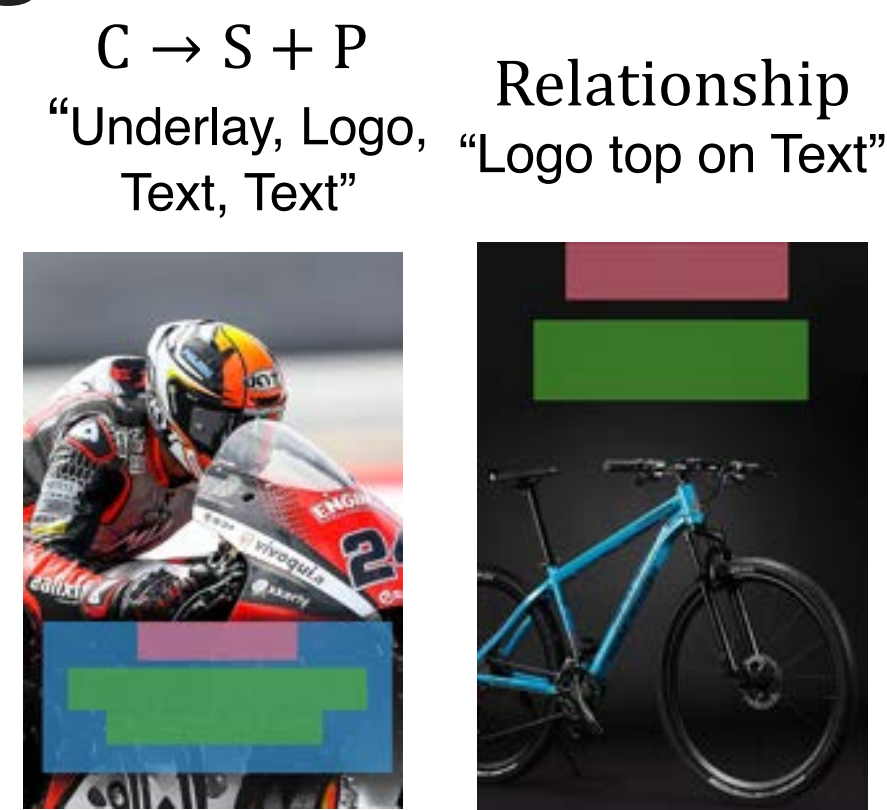


Retrieval-augmented content-aware layout generation

- Retrieve nearest neighbor examples based on the input image and use them as a reference to augment the generation process.
- Propose a *Retrieval-Augmented Layout Transformer (RALF)* designed to integrate retrieval augmentation for *content-aware layout generation*.

Challenges

- Data Scarcity & Training Efficiency
- High-dimensional Layout Structure
- Content-Layout Harmonization
- Controllability to User-Specified Constraints



Constrained generation tasks

Contributions

- Demonstrate the effectiveness of retrieval augmentation** in content-aware layout generation.
- Propose RALF:** Retrieval-Augmented Generation + Autoregressive Transformer.
- Show that RALF successfully generates high-quality layouts,** significantly outperforming baselines.

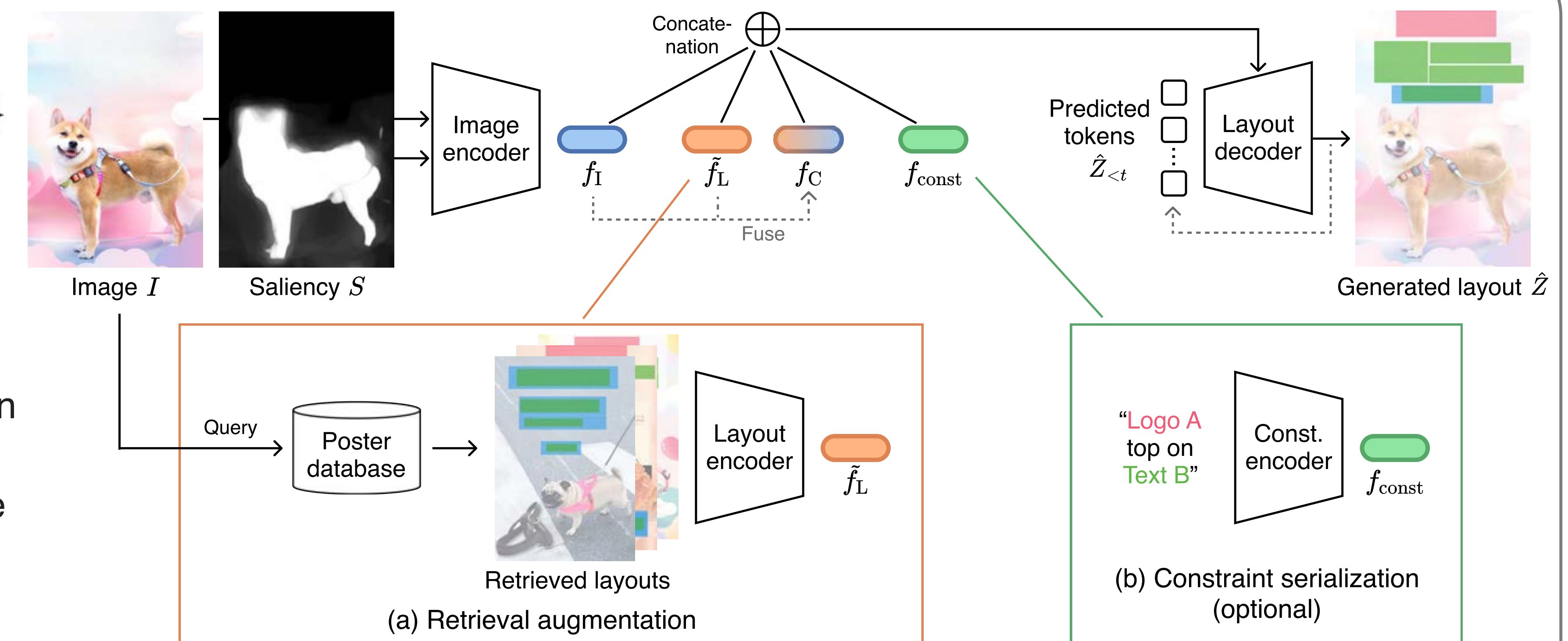
Preliminaries

- Layout: $L = \{l_1, \dots, l_T\} = \{(c_1, \mathbf{b}_1), \dots, (c_T, \mathbf{b}_T)\}$
 bounding box: $\mathbf{b} \in [0, 1]^4$, category: $c_i \in \{1, \dots, C\}$
- Flattened 1D sequence of layout:
 $Z = (bos, c_1, x_1, y_1, \dots, w_T, h_T, eos) \in \mathbb{N}^{5T+2}$

How does RALF generate layouts?

- Encodes** input canvas image and saliency map.
- Retrieves** nearest neighbor layout examples based on the input image.
- Fuses** the features of retrieved layouts with the image feature using cross-attention.
- Incorporates** user-specified constraints.
- Autoregressively generates** a layout.

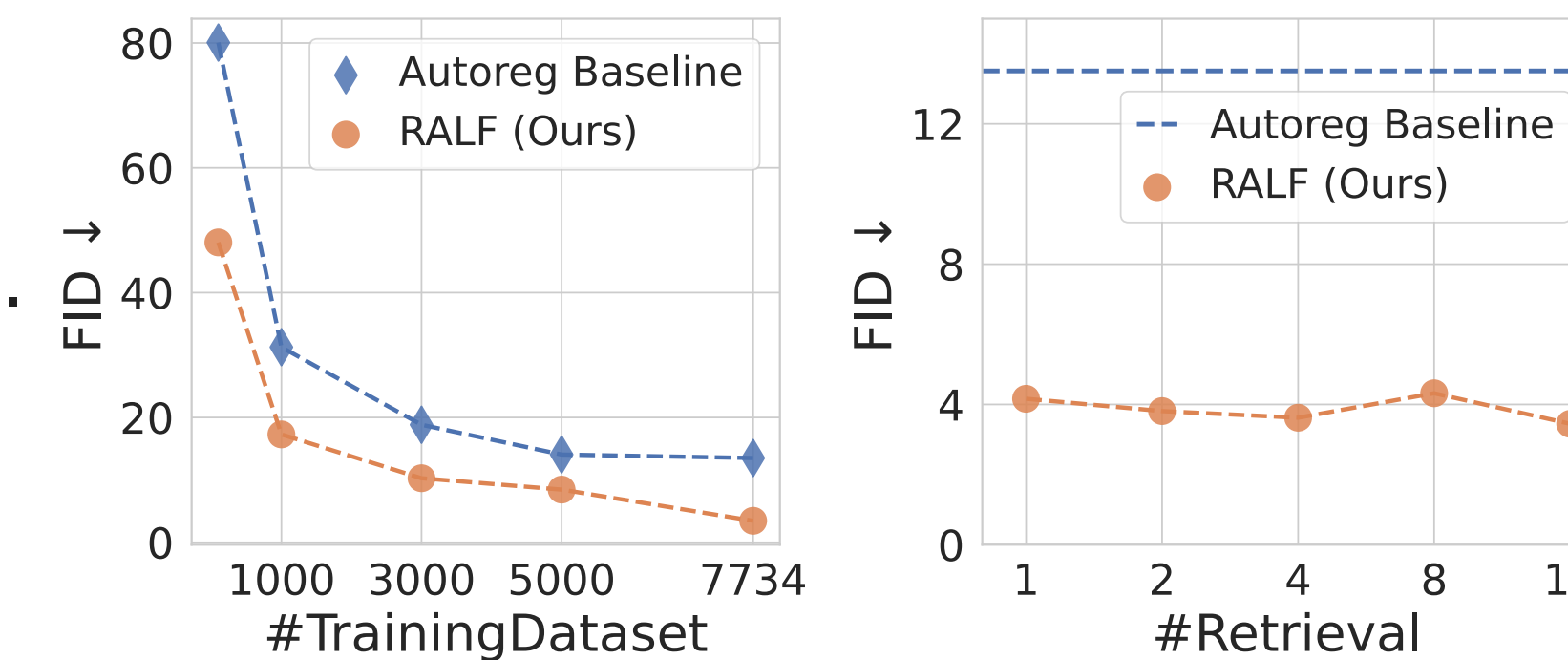
Method



Analysis

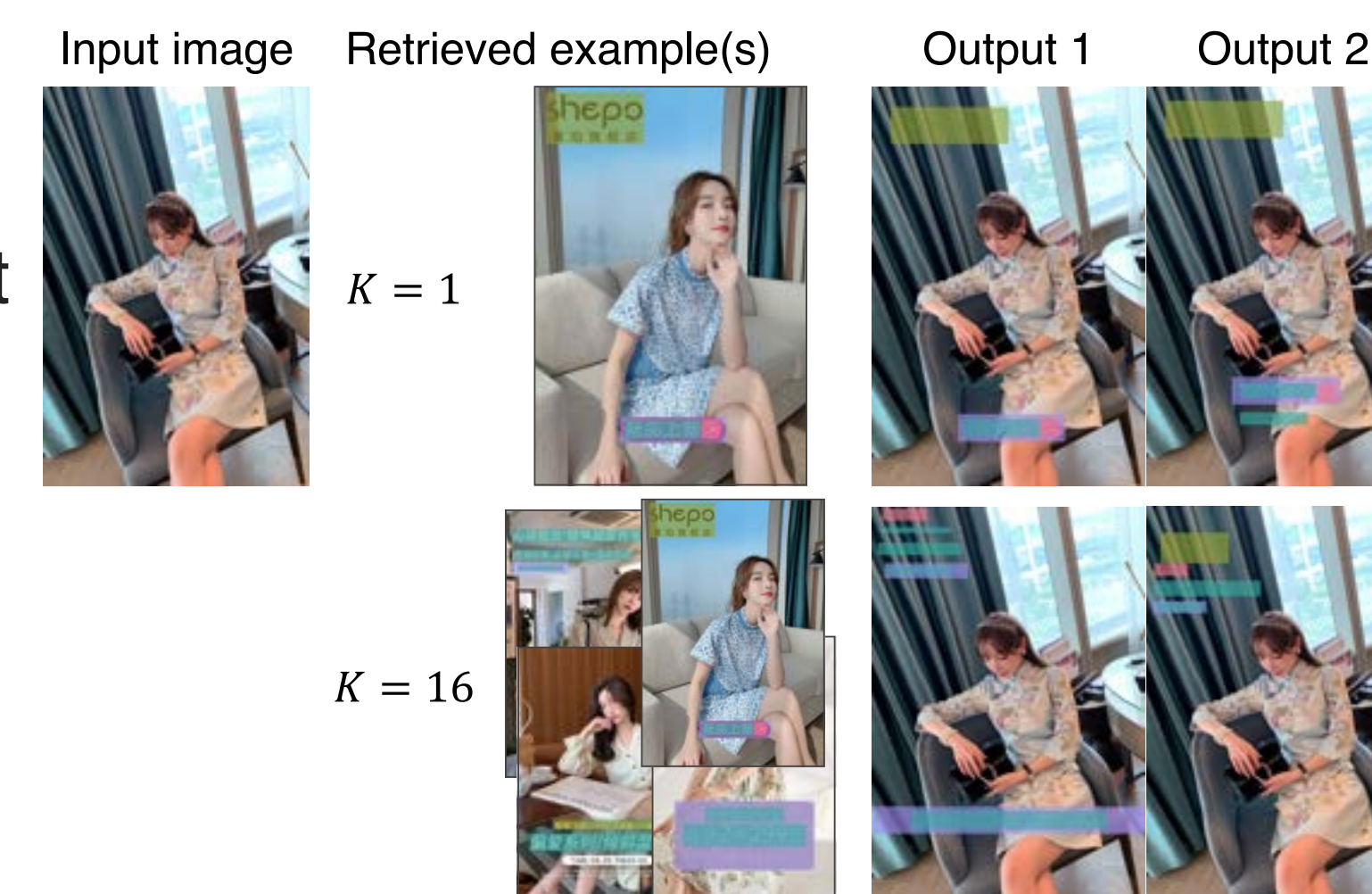
How effective RALF?

- Effective** regardless of the training dataset size.
- Not sensitive** to the number of retrieved layouts.



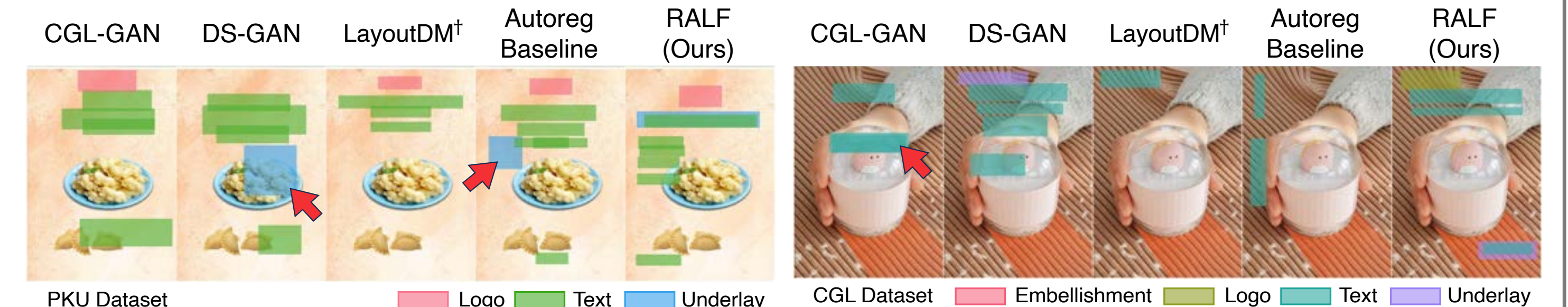
How different K affects the generated results?

- $K = 1$: the generated layout is similar to the reference.
- $K = 16$: a variety of layouts are generated.



Experiments

Qualitative Results



Quantitative Results

Method	#Params	PKU				
		Content		Graphic		
		Occ ↓	Rea ↓	Und ↑	Ove ↓	FID ↓
Real Data	-	0.112	0.0102	0.99	0.0009	1.58
Top-1 Retrieval	-	0.212	0.0218	0.99	0.002	1.43
CGL-GAN [53]	41M	0.138	0.0164	0.41	0.074	34.51
DS-GAN [18]	30M	0.142	0.0169	0.63	0.027	11.80
ICVT [7]	50M	0.146	0.0185	0.49	0.318	39.13
LayoutDM [†] [19]	43M	0.150	0.0192	0.41	0.190	27.09
Autoreg Baseline	41M	0.134	0.0164	0.43	0.019	13.59
RALF (Ours)	43M	0.119	0.0128	0.92	0.008	3.45

Method	Retrieval	Occ ↓	Rea ↓	Und ↑	Ove ↓	FID ↓
CGL-GAN		0.138	0.0164	0.41	0.074	34.51
CGL-GAN	✓	0.144	0.0164	0.63	0.039	13.28
LayoutDM [†]		0.150	0.0192	0.41	0.190	27.09
LayoutDM [†]	✓	0.123	0.0144	0.51	0.091	10.03

Retrieval augmentation for other methods

Train	Test	Method	Occ ↓	Rea ↓	Und ↑	Ove ↓
CGL	PKU	Autoreg Baseline	0.176	0.0276	0.84	0.037
		RALF (Ours)	0.144	0.0249	0.96	0.023
PKU	CGL	Autoreg Baseline	0.341	0.0464	0.29	0.037
		RALF (Ours)	0.286	0.0355	0.79	0.036

Unconstrained generation results on PKU dataset

Out-of-domain generalization