大阪大学 THE UNIVERSITY OF OSAKA

LINEヤフー

VASCAR: Content-Aware Layout Generation via Visual-Aware Self-Correction

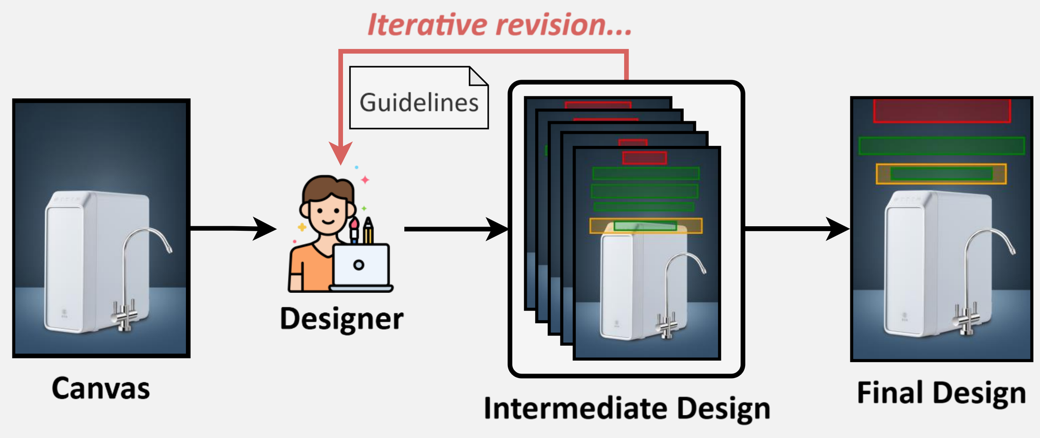
Jiahao Zhang♣, Ryota Yoshihashi♢, Shunsuke Kitada♢, Atsuki Osanai♢, Yuta Nakashima♣

♣The University of Osaka

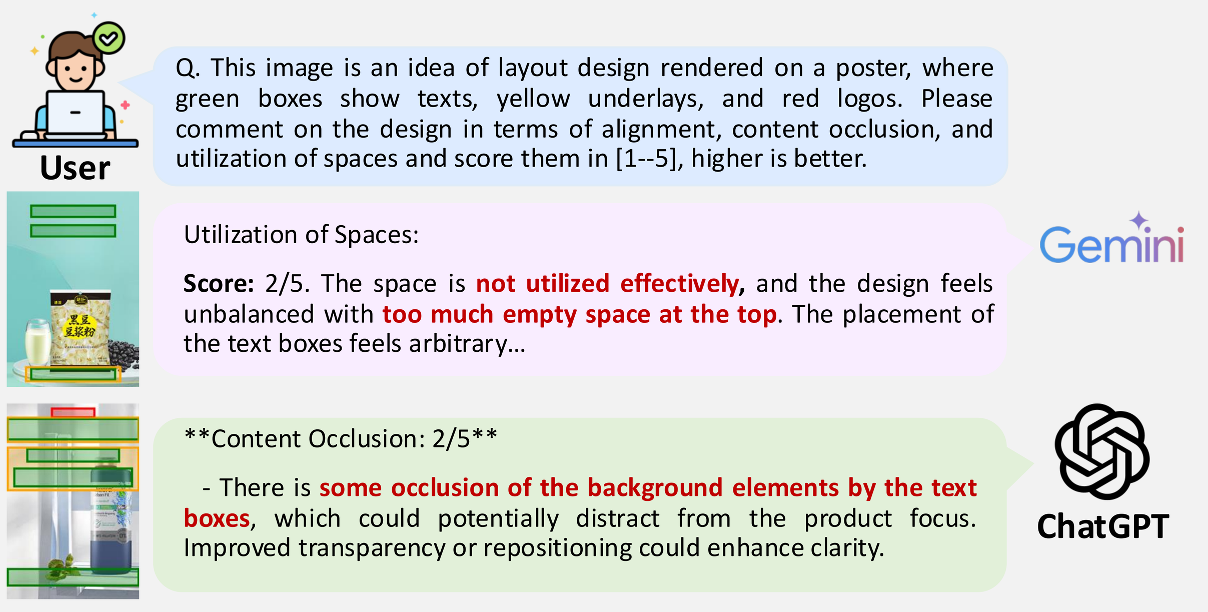
♢LY Corporation

OS3A-09

Motivation



(a) The iterative revision workflow of designers.



(b) Inspiring conversations with Gemini/ChatGPT.

Introduction

What is the Content-aware layout generation task?

- Content-aware layout generation aims to automatically arrange visual elements (e.g., text, logo, underlay) on a canvas based on its visual content — a task essential in applications such as poster and magazine design.
  While recent **large language models (LLMs)** can generate structured layouts via HTML or JSON, they lack the ability to *see*, limiting their effectiveness when visual cues are crucial.

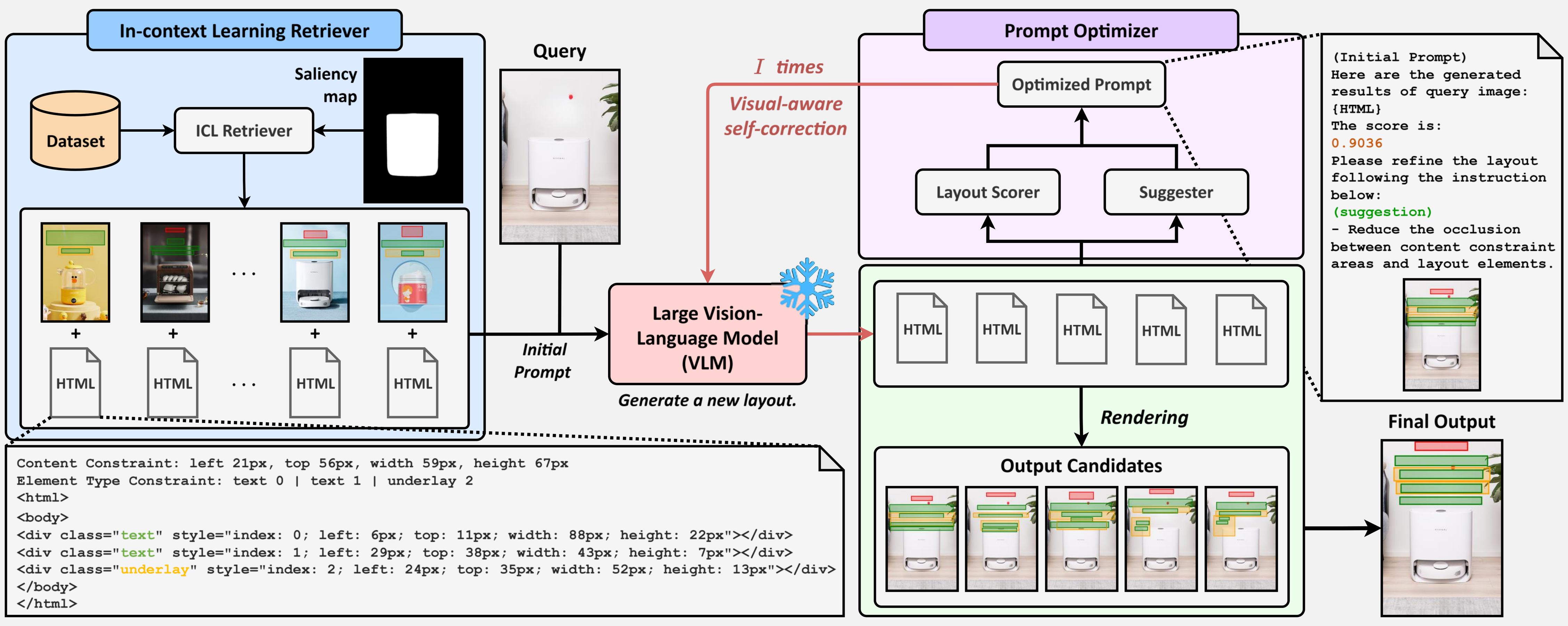
Key Challenges

- **Lack of visual understanding** in LLMs hinders layout quality.
- **High cost of training** generative models on layout data.
- **Absence of iterative refinement**, which is common in human design workflows.

Our Solution: VASCAR

- We introduce ***VASCAR***, a **training-free framework** for content-aware layout generation using **large vision-Language models (VLMs)** such as **GPT-4o** and **Gemini**.
- Inspired by how human designers work, ***VASCAR***:
  - **Generates layout candidates** via few-shot in-context learning (ICL).
  - **Evaluates layout quality** using automatic multi-criteria scoring.
  - **Refines outputs iteratively** based on visual feedback and textual suggestions.

Method



The overview of proposed ***VASCAR***.

ICL Retriever:

Retrieves a small set of **in-context learning (ICL) examples** similar to the **Query** canvas, using **saliency map similarity**.

$$s(x_{test}, x_j) \triangleq \mathrm{IoU}(m_{test}, m_j) = \frac{m_{test} \cap m_j}{m_{test} \cup m_j}$$

x_{test}, x_j are **Query** and ICL samples, m_{test}, m_j are corresponding saliency maps.

Each exemplar includes:

- Rendered layout (colored bounding boxes)
- ***HTML format*** layout description

Layout Generator:

Generates initial layout candidates using a **frozen VLM** given the query canvas and ICL examples.

$$Y_q^0 = G(x_q, \emptyset; p_0)$$

- x_q: **Query** canvas
- p_0: Initial prompt
- $G(\cdot)$: VLM generator (GPT-4o & Gemini)
- Y_q^0: Set of initial layout candidates

Output is in ***HTML format***, describing bounding box coordinates and element types.

Visual-Aware Self-Correction:

Performs **iterative layout refinement** using visual feedback from rendered images and evaluation scores.

Refinement formula (iteration i):

$$Y_q^i = G(x_q, Y_q^{i-1}; p(Y_q^{i-1}))$$

- Renders previous layouts onto canvas
- Scores each candidate
- Adds suggestions as text prompt
- Feeds it back to the VLM

Prompt Optimizer:

Constructs the next-step **multi-modal prompt** for the VLM.

Input:

- **Query** canvas image
- Top-k previous rendered layouts
- ICL examples
- Suggestions + scores

Layout Scorer:

Assigns a **fused quality score** to each layout candidate based on multiple normalized criteria.

$$v(y_q) = \sum_{m \in \mathcal{M}} \lambda_m \cdot f_m(y_q)$$

- $\mathcal{M}$: Set of evaluation metrics
- $f_m(y_q)$: Normalized score for metric m
- λ_m: Weight for each metric

Suggester:

Provides **natural language suggestions** to refine the layout, based on low-scoring metrics, threshold θ_m is calculated by ICL examples:

$$\theta_m = \frac{1}{|S(x_q)|} \sum_{y \in S(x_q)} f_m(y)$$

- If $f_m(y_q) < \theta_m$, add a suggestion related to that score, such as:

  "Reduce the overlap between the text and image."

Results

Experimental Setup:

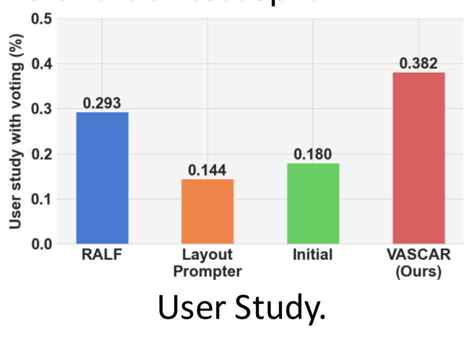
- **Datasets:**
  We conducted our experiments on two opensource datasets: PKU and CGL, both of which contain e-commerce posters featuring shopping product images.
- **Compared Methods:**
  - **LLM-based:**
    LayoutPrompter, PosterLlama
  - **Generative baselines (trained):**
    CGL-GAN, DS-GAN, ICVT, LayoutDM, Autoreg, RALF
- **Evaluation Metrics:**
  - **Content metrics:** Occusion, Unreadability
  - **Graphic metrics:** Overlay, Non-alignment, Underlay Effectiveness, FID
- **Implementation Details.**
  - gemini-1.5-flash and gpt-4o are used as VLMs.
  - ICL examples: 10.
  - Self-correction: 15 for Gemini, 5 for GPT-4o

Method	Training-free	PKU Content Occ ↓	PKU Content Rea ↓	PKU Graphic Align ↓	PKU Graphic Und ↑	PKU Graphic Ove ↓	PKU Graphic FID↓	CGL Content Occ ↓	CGL Content Rea ↓	CGL Graphic Align ↓	CGL Graphic Und ↑	CGL Graphic Ove ↓	CGL Graphic FID↓
Real Data	-	0.112	0.0102	0.0038	0.99	0.0009	1.58	0.125	0.0170	0.0024	0.98	0.0002	0.79
CGL-GAN [68]	✗	0.138	0.0164	0.0031	0.41	0.0740	34.51	0.157	0.0237	0.0032	0.29	0.1610	66.75
DS-GAN [20]	✗	0.142	0.0169	0.0035	0.63	0.0270	11.80	0.141	0.0229	0.0026	0.45	0.0570	41.57
ICVT [5]	✗	0.146	0.0185	0.0023	0.49	0.3180	39.13	**0.124**	0.0205	0.0032	0.42	0.3100	65.34
LayoutDM [22]	✗	0.150	0.0192	0.0030	0.41	0.1900	27.09	0.127	0.0192	0.0024	0.82	0.0200	2.36
Autoreg [19]	✗	0.134	0.0164	0.0019	0.43	0.0190	13.59	0.125	0.0190	0.0023	0.92	0.0110	2.89
RALF [19]	✗	0.119	0.0128	0.0027	0.92	0.0080	3.45	0.125	0.0180	0.0024	0.98	0.0040	**1.32**
PosterLlama† [49]	✗	–	–	–	–	–	–	0.154	0.0135	0.0008	0.97	0.0030	2.21
LayoutPrompter† [33]	✓	0.220	0.0169	0.0006	0.91	0.0003	3.42	0.251	0.0179	**0.0004**	0.89	**0.0002**	4.59
VASCAR (GPT-4o)	✓	0.129	**0.0091**	**0.0002**	**0.99**	**0.0002**	**3.14**	0.141	**0.0102**	0.0005	**0.99**	**0.0002**	5.69
VASCAR (Gemini)	✓	**0.113**	0.0117	0.0013	0.98	0.0003	3.34	0.125	0.0122	0.0010	0.98	0.0007	6.27

Unconstrained generation results on the PKU and CGL test split.

Method	PKU unannotated Occ ↓	PKU unannotated Rea ↓	PKU unannotated Und ↑	PKU unannotated Ove ↓	CGL unannotated Occ ↓	CGL unannotated Rea ↓	CGL unannotated Und ↑	CGL unannotated Ove ↓
CGL-GAN [68]	0.191	0.0312	0.32	0.0690	0.481	0.0568	0.26	0.2690
DS-GAN [20]	0.180	0.0301	0.52	0.0260	0.435	0.0563	0.29	0.0710
ICVT [5]	0.189	0.0317	0.48	0.2920	0.446	0.0425	0.67	0.3010
LayoutDM [22]	0.165	0.0285	0.38	0.2010	0.421	0.0506	0.49	0.0690
Autoreg [19]	0.154	0.0274	0.35	0.0220	0.384	0.0427	0.76	0.0580
RALF [19]	0.133	0.0231	0.87	0.0180	0.336	0.0397	0.93	0.0270
VASCAR (Ours)	**0.132**	**0.0151**	**0.98**	**0.0002**	**0.289**	**0.0300**	**0.98**	**0.0007**

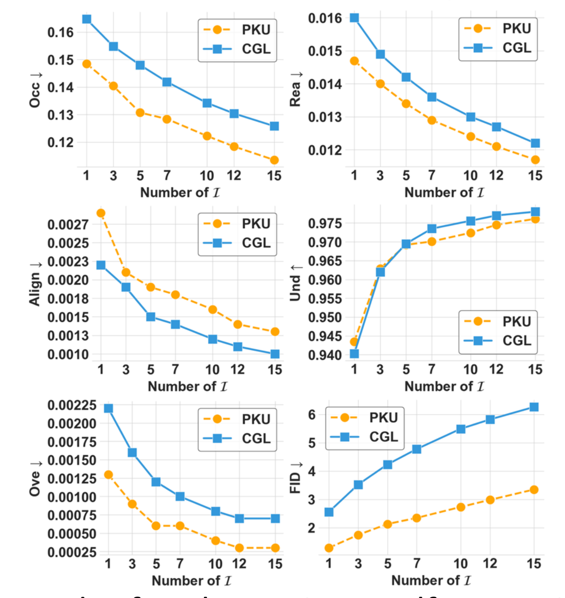
Unconstrained generation results on the unannotated test split.



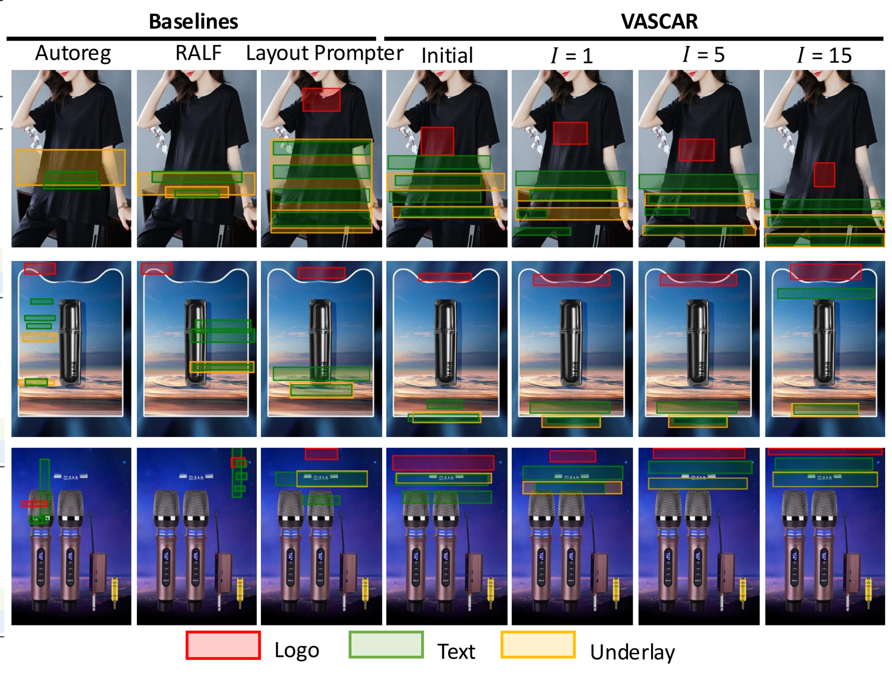
User Study.

Method	PKU Content Occ ↓	PKU Content Rea ↓	PKU Graphic Und ↑	PKU Graphic Ove ↓	PKU Graphic FID↓	CGL Content Occ ↓	CGL Content Rea ↓	CGL Graphic Und ↑	CGL Graphic Ove ↓	CGL Graphic FID↓
Real Data	0.112	0.0102	0.99	0.0009	1.58	0.125	0.0170	0.98	0.0002	0.79
C → S + P										
CGL-GAN	0.132	0.0158	0.48	0.0380	11.47	0.140	0.0213	0.65	0.0470	23.93
LayoutDM	0.152	0.0201	0.46	0.1720	20.56	0.127	0.0192	0.79	0.0260	3.39
Autoreg	0.135	0.0167	0.43	0.0280	10.48	0.124	0.0188	0.89	0.0150	1.36
RALF	0.124	0.0138	0.90	0.0100	**2.21**	0.126	0.0180	0.97	0.0060	**0.50**
VASCAR (GPT-4o)	0.117	**0.0094**	**1.00**	**0.0002**	3.01	0.139	**0.0099**	**0.99**	**0.0002**	4.86
VASCAR (Gemini)	**0.107**	0.0100	0.99	0.0003	2.82	**0.123**	0.0111	0.98	0.0005	5.51
C + S → P										
CGL-GAN	0.129	0.0155	0.48	0.0430	9.11	0.129	0.0202	0.75	0.0270	6.96
LayoutDM	0.143	0.0185	0.45	0.1220	24.90	0.127	0.0190	0.82	0.0210	2.18
Autoreg	0.137	0.0169	0.46	0.0280	5.46	0.127	0.0191	0.88	0.0130	0.47
RALF	0.125	0.0138	0.87	0.0100	**0.62**	0.128	0.0185	**0.96**	0.0060	**0.21**
VASCAR (GPT-4o)	0.123	0.0117	**0.90**	**0.0009**	1.11	0.140	0.0132	0.88	**0.0003**	1.67
VASCAR (Gemini)	**0.104**	**0.0107**	0.88	0.0018	2.21	**0.122**	**0.0123**	0.85	0.0028	2.39
Completion										
CGL-GAN	0.150	0.0174	0.43	0.0610	25.67	0.174	0.0231	0.21	0.1820	78.44
LayoutDM	0.135	0.0175	0.35	0.1340	21.70	0.127	0.0192	0.76	0.0200	3.19
Autoreg	0.125	0.0161	0.42	0.0230	5.96	**0.124**	0.0185	0.91	0.0110	2.33
RALF	0.120	0.0140	0.88	0.0120	**1.58**	0.126	0.0185	0.96	0.0050	**1.04**
VASCAR (GPT-4o)	0.120	**0.0063**	**1.00**	**0.0004**	6.19	0.137	**0.0068**	**0.99**	**0.0005**	5.57
VASCAR (Gemini)	**0.119**	0.0097	0.99	0.0006	4.74	0.135	0.0098	0.98	0.0009	5.18
Refinement										
CGL-GAN	0.122	0.0141	0.39	0.0900	6.40	0.124	0.0182	0.86	0.0240	1.20
LayoutDM	0.115	0.0174	0.40	0.0630	2.86	0.127	0.0196	0.75	0.0150	1.98
Autoreg	0.131	0.0171	0.40	0.0220	5.89	0.126	0.0183	0.89	0.0040	0.15
RALF	0.113	0.0109	0.95	0.0040	**0.13**	0.126	0.0176	**0.98**	0.0020	**0.14**
VASCAR (GPT-4o)	0.108	**0.0072**	**0.99**	**0.0005**	1.18	0.125	**0.0091**	0.95	**0.0008**	0.59
VASCAR (Gemini)	**0.104**	0.0095	0.97	0.0010	0.32	**0.114**	0.0096	0.96	0.0013	0.37
Relationship										
Autoreg	0.140	0.0177	0.44	0.0280	10.61	0.127	0.0189	0.88	0.0150	1.28
RALF	0.122	0.0141	0.85	0.0090	2.23	**0.126**	0.0184	**0.95**	0.0060	**0.55**
VASCAR (GPT-4o)	0.151	0.0139	0.92	0.0011	**1.94**	0.153	0.0139	0.91	**0.0004**	2.61
VASCAR (Gemini)	**0.119**	**0.0117**	**0.96**	**0.0008**	2.00	0.132	**0.0123**	**0.95**	0.0011	3.72

Quantitative result of five constrained generation tasks on the PKU and CGL test splits.



The trends of each metric on self-correction I.



Visual comparison of baselines and VASCAR with different values of I.

Setting	Content Occ ↓	Content Rea ↓	Graphic Align ↓	Graphic Und ↑	Graphic Ove ↓	Graphic FID↓
Rendered Image (Ours)	**0.1304**	**0.0134**	**0.0017**	0.97	0.0012	2.13
Text-only	0.1529	0.0153	0.0024	0.97	0.0009	1.91
Saliency Map	0.1356	0.0140	0.0023	0.97	0.0009	**1.69**
Inpainting Image	0.1312	0.0137	0.0022	**0.98**	**0.0004**	2.37
Original Poster	0.1315	**0.0134**	0.0023	**0.98**	0.0006	2.34

Multi-modal analysis for VASCAR on the PKU test split.

Setting	Content Occ ↓	Content Rea ↓	Graphic Align ↓	Graphic Und ↑	Graphic Ove ↓	Graphic FID↓
Number of ICL Examples (M)						
1	0.2014	0.0184	0.0026	0.84	0.0017	**1.45**
3	0.1617	0.0160	0.0020	0.90	0.0016	2.27
5	0.1466	0.0150	0.0018	0.95	**0.0006**	2.25
10 (Ours)	**0.1304**	**0.0134**	**0.0017**	**0.97**	0.0012	2.13
Number of Output Candidates ($\lvert Y_q \rvert$)						
1	0.1659	0.0180	0.0023	0.89	0.0011	2.29
3	0.1393	0.0150	**0.0017**	0.96	0.0010	2.24
5 (Ours)	0.1304	0.0134	**0.0017**	**0.97**	0.0012	2.13
10	**0.1225**	**0.0127**	0.0025	**0.97**	**0.0008**	**2.04**
Ablation Study on λ						
VASCAR (Ours)	0.1304	0.0134	**0.0017**	**0.97**	0.0012	2.13
Initial	0.2080	0.0218	0.0040	0.97	0.0013	1.05
w/o Occ	0.1709	**0.0114**	0.0022	**0.97**	**0.0002**	**1.95**
w/o Rea	0.1284	0.0162	0.0024	**0.97**	0.0008	2.03
w/o Ove	0.1301	0.0137	0.0018	**0.97**	0.0030	2.23
w/o Und	**0.1237**	0.0130	0.0023	0.68	0.0007	2.14

Comparison of results across various settings on PKU.