

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

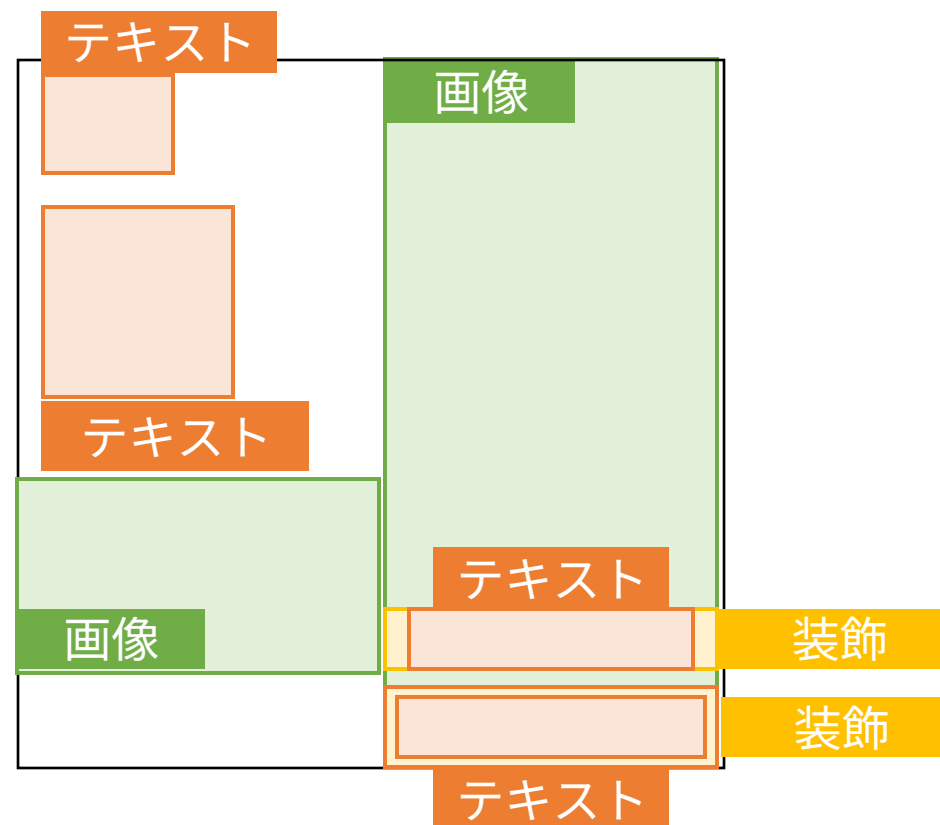
画像の認識・理解シンポジウム MIRU2025
張 家豪♠, 吉橋 亮太♦, 北田 俊輔♦, 長内 淳樹♦, 中島 悠太♠
♠大阪大学
♦LINEヤフー

レイアウトとは

- 適切に配置されたグラフィック要素の集合からなるレイアウトは、グラフィックデザインにおいて重要な役割を果たしている



Yahoo Shopping画像



レイアウト生成とは

□ どの要素を、どこに大きさを配置するかきめるタスク

制約条件ありの生成

種類を条件とした生成

toolbar

icon

種類とサイズを条件とした生成

Toolbar(90, 12)

Icon (12,12)

関係性を条件とした生成

toolbar

icon

toolbar on top

icon on bottom

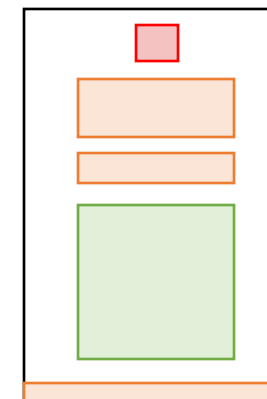
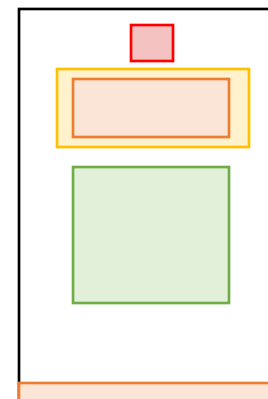
レイアウト完成

Toolbar (0, 5, 90, 12)

レイアウト改善

Toolbar (2, 8, 90, 12)

Content-aware
Layout
Generation



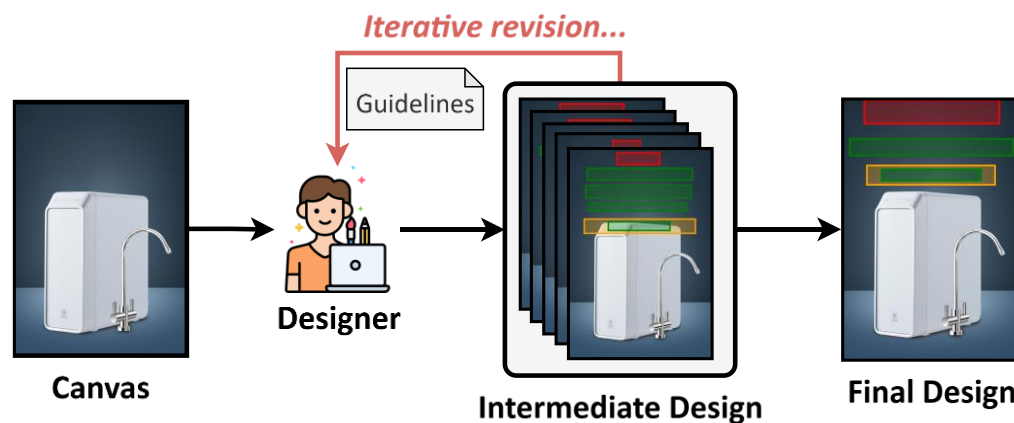
レイアウト生成の先行研究

□ 先行研究は有望な結果を達成していますが、いくつかの制約がある

- すべてのレイアウト生成タスクに同時に対処することが難しく、特定のタスクに特化して調整されているため、他のタスクには適用できない点が課題
- 既存の手法はデータ効率が良くない、モデルのトレーニングに大量のレイアウトデータを必要とする
- パラメータ効率が良くない。美しいレイアウトを生成するには、LLMのファインチューニングが必要

モチベーション

- デザイナーがガイドラインを基に行う、繰り返し修正のワークフロー
- LVLMがレンダリング画像 (bounding box) のレイアウトを評価できる



User

Q. This image is an idea of layout design rendered on a poster, where green boxes show texts, yellow underlays, and red logos. Please comment on the design in terms of alignment, content occlusion, and utilization of spaces and score them in [1--5], higher is better.

Utilization of Spaces:

Score: 2/5. The space is **not utilized effectively**, and the design feels unbalanced with **too much empty space at the top**. The placement of the text boxes feels arbitrary...

****Content Occlusion: 2/5****

- There is **some occlusion of the background elements by the text boxes**, which could potentially distract from the product focus. Improved transparency or repositioning could enhance clarity.

Gemini

ChatGPT

VASCAR

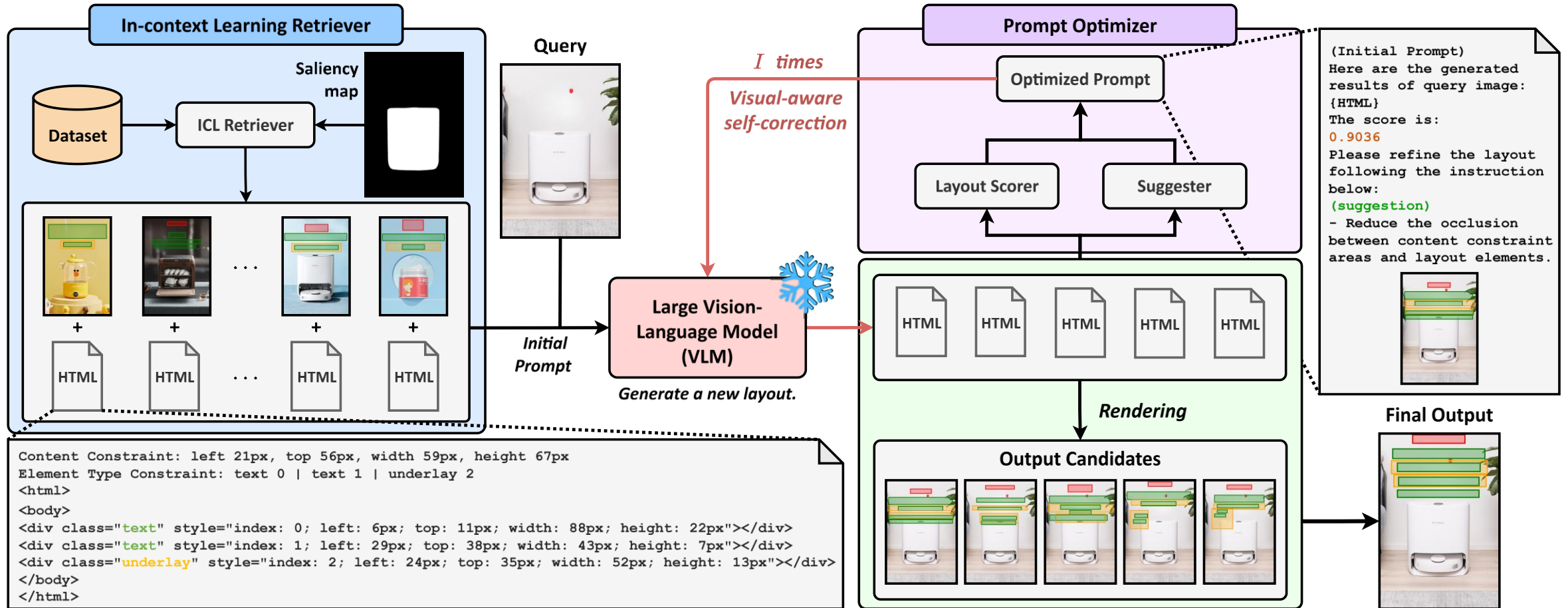
- 大規模画像-言語モデル (LVLM) を活用したContent-awareレイアウト生成手法である **V**isual-**A**ware **S**elf-**C**orrection **L**Ayout Gene**R**ation (**VASCAR**) を提案した
 - **マルチモーダル**：レンダリング画像とHTML形式のレイアウト入力を使用して、モデルを指示する
 - **Training-free and In-context Learning**: few-shotサンプルを使用してモデルを指示し、モデルのトレーニングが必要なし
 - **Self-correction**: 評価指標を用いてモデルに修正結果を指示し、レイアウト生成における有用な指示を組み込む手法を提案した



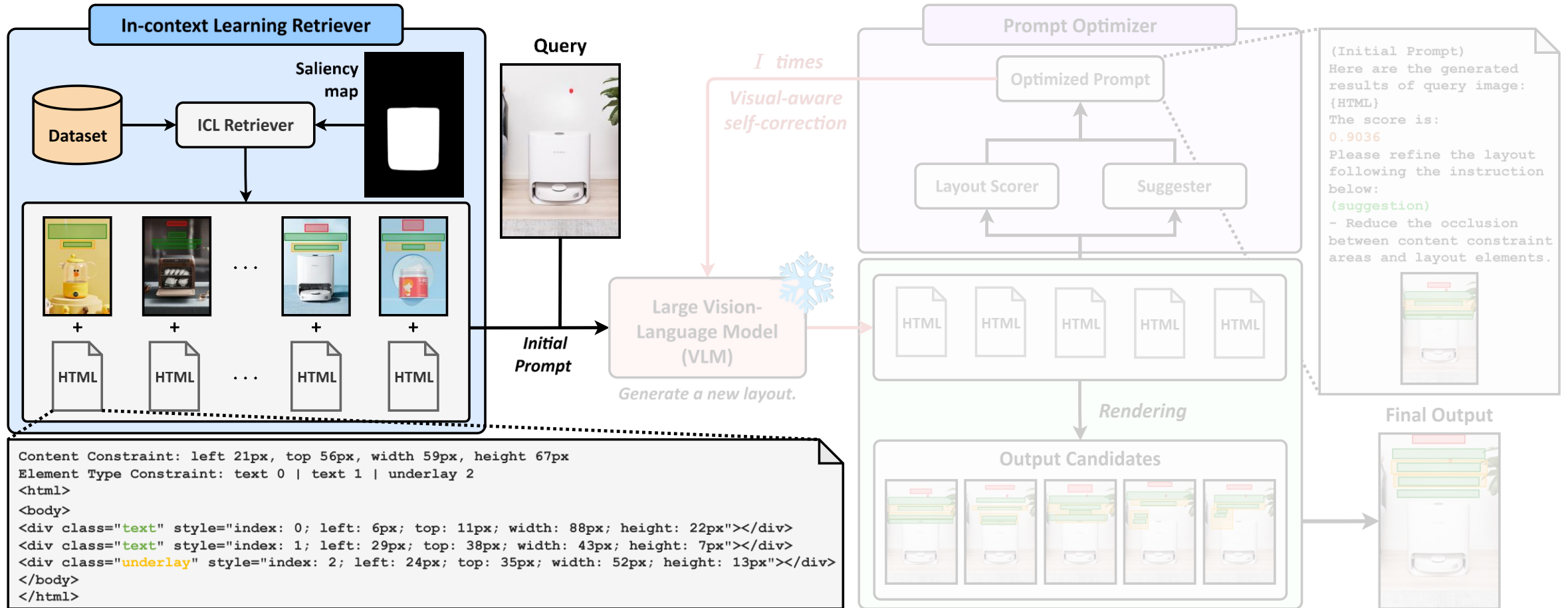
Method

- Overview
- ICL Retriever
- Layout Generator
- Self-Correction

Overview



ICL Retriever



ICL Retriever (ICLサンプルの選択)

- Content-aware Layout Generationとは、saliency mapの領域でレイアウトを生成しないようなタスクである
- モデルを指示するために、クエリサンプルと類似したサンプルの選択が必要
- Intersection over Union (IoU) に基づくICLサンプルを選択する：

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

x_{test}, x_j はクエリ画像とICLサンプル、 m_{test}, m_j は対応するsaliency map

ICL Retriever (ICLサンプルの入力)

- **Content Constraint** の情報がテキスト中に付与し、ICLサンプルのレンダリング画像をモデルに入力する

Input Image 1:

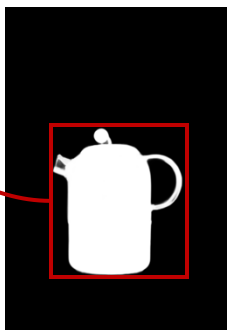
Content Constraint: left 21px, top 56px, width 59px, height 67px

Element Type Constraint: text 0 | text 1 | underlay 2

```
<html>
<body>
<div class="text" style="index: 0; left: 6px; top: 11px; width: 88px; height: 22px"></div>
<div class="text" style="index: 1; left: 29px; top: 38px; width: 43px; height: 7px"></div>
<div class="underlay" style="index: 2; left: 24px; top: 35px; width: 52px; height: 13px"></div>
</body>
</html>
```



Input Image 1

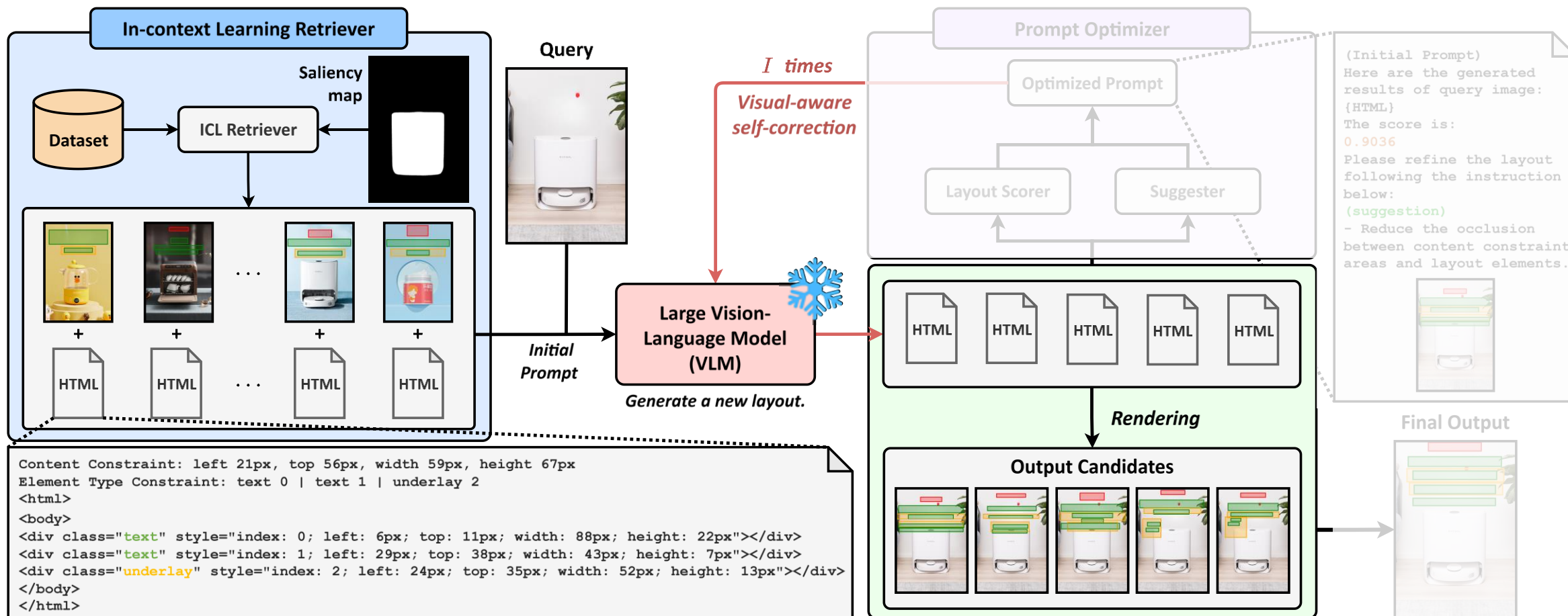


Saliency

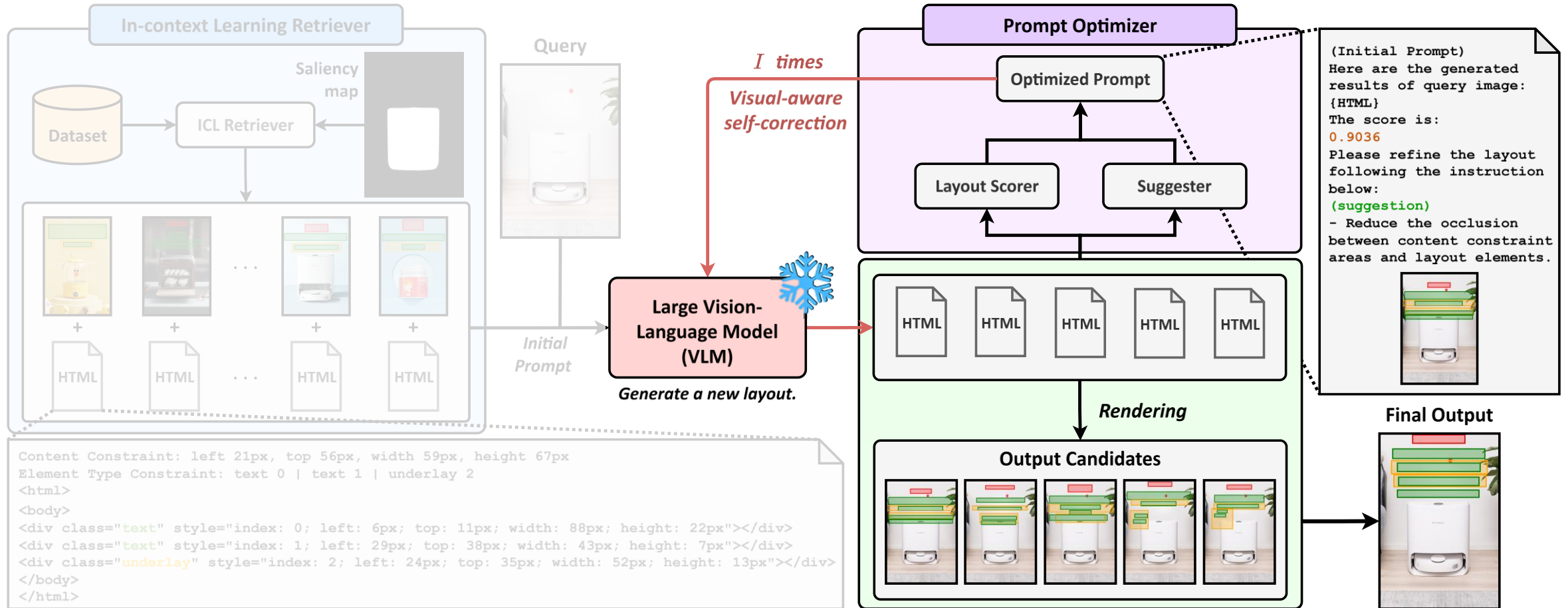


Canvas

Layout Generator



Self-Correction



Layout Scorer

- レイアウト評価の各指標を用いて生成結果を評価し、これらの指標を0～1のスコアに統合して、重みを使用してそれぞれを調整する
- 評価指標は以下の通り
 - Overlap (0.1)：各要素の overlap を評価
 - Underlay effectiveness (0.1)：underlayが他の要素を装飾する程度を評価
 - Occlusion (0.4)：要素がコンテンツ（Saliency map）を覆う程度を評価
 - Non-Readability (0.4)：テキスト要素の可読性を評価

Suggester

- Layout scorerから得られたスコアだけを用いてモデルを指示するのは最適な方法ではない。そのため、スコアに基づいてモデルに具体的な調整を行う手法を提案した
- 生成結果の特定の指標が設定された閾値を下回った場合（閾値はground truthを含むICLサンプルに基づいて各指標の平均値で設定される）、具体的な指示を行う**提案 (Suggestion)**を加える



Overlap ↓ : 0.0 (0.0)
Alignment ↓ : 2.8868 (3.6203)
Underlay ↑ : 1.0 (1.0)
Utility ↑ : 0.09636 (0.3040)
Occlusion ↓ : 0.8372 (0.1331)
Readability ↓ : 0.0 (0.0088)

Here are the generated results of query image: Red is Logo; Green is Text; Yellow is Underlay.

{HTML}

The score is:
0.3631

Please refine the layout following the instruction below:
(Suggestion)

- Reduce the overlap between content constraint areas and layout elements.

Optimized Prompt

- ❑ Optimized promptには、5つの生成結果のスコアと提案、およびレンダリングされた画像が含まれる。このoptimized promptをinitial promptに追加し、再びモデルに指示を与えて新しい結果を生成する
- ❑ Self-correctionは 1 回繰り返され、各ステップでTop-5の結果optimized promptのサンプルとして使用し、再びモデルに指示を与える

```
(Initial Prompt)
Here are the generated results of query image:

{HTML}

The score is:
0.9036

Please refine the layout following the instruction below:
(suggestion)
- Reduce the occlusion between content constraint areas and
  layout elements.
.....
```



Experiment Setup

□ PKUとCGLデータセットで実験した

□ VASCAR:

- LVLM: Gemini-1.5-Flash, GPT-4o
- ICL Exampleの数 : 10
- Self-correctionの数 (モデルを回す回数) : 15 (Gemini-1.5-Flash), 5 (GPT-4o)
- Output Candidatesの数 : 5

□ 評価指標 :

■ Content

- Occ ↓ : 要素がコンテンツ (Saliency map) を覆うIoUを評価
- Rea ↓ : テキスト要素の不可読性を評価

■ Graphic

- Ove ↓ : 各要素の overlap を評価
- Align ↓ : 各要素の整列状況の評価
- Und ↑ : underlayが完全に他の要素を装飾する平均値
- FID ↓ : レイアウトとGround Truthのレイアウトとの分布を比較

Results

- VASCARは、PKUおよびCGLデータセットにおいて、ほぼすべての指標で最適の結果を達成した

Method	Training-free	PKU						CGL					
		Content		Graphic				Content		Graphic			
		Occ ↓	Rea ↓	Align ↓	Und ↑	Ove ↓	FID ↓	Occ ↓	Rea ↓	Align ↓	Und ↑	Ove ↓	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	<u>0.125</u>	0.0190	0.0023	0.92	0.0110	2.89
RALF [19]	✗	<u>0.119</u>	0.0128	0.0027	0.92	0.0080	3.45	<u>0.125</u>	0.0180	0.0024	<u>0.98</u>	0.0040	1.32
PosterLlama [†] [49]	✗	—	—	—	—	—	—	0.154	0.0135	0.0008	0.97	0.0030	<u>2.21</u>
LayoutPrompter [†] [33]	✓	0.220	0.0169	<u>0.0006</u>	0.91	<u>0.0003</u>	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	<u>0.0005</u>	0.99	0.0002	5.69
VASCAR (Gemini)	✓	0.113	<u>0.0117</u>	0.0013	<u>0.98</u>	<u>0.0003</u>	<u>3.34</u>	<u>0.125</u>	<u>0.0122</u>	0.0010	<u>0.98</u>	<u>0.0007</u>	6.27