

From Elements to Design: A Layered Approach for Automatic Graphic Design Composition

Supplementary Material

A. Implementation Details of Layer Planning

Layer planning is achieved by predicting semantic labels, including background, underlay, logo/image, text, and embellishment, for input elements. In the implementation, we can easily identify text elements since they have the text content attribute. For non-text elements, we then leverage GPT-4o [1] to categorize them into the other four labels in a zero-shot manner. We carefully design the following *Element Labeling Prompt* to accomplish this effectively.

Element Labeling Prompt

You are an excellent graphic designer. Your task is to determine the role of the given element, which is rendered as an image. There are 4 possible options: **Background**, **Underlay**, **Logo/Image** or **Embellishment**. Please refer to the detailed descriptions below to make your prediction.

Background: The foundational layer of the design, typically large in size and covering the entire canvas. It may consist of a solid color, gradient, landscape image, or similar visual foundation.

Underlay: A supportive layer placed beneath key content, often used to create contrast or highlight the main design elements, such as borders, buttons, color overlays, and so on.

Logo/Image: A core visual element that represents a brand, product, or entity. It combines both imagery and logo elements to capture attention and convey the primary message.

Embellishment: Decorative elements that enhance visual appeal without conveying core information. These elements add style to the design. Note that they are usually small in size.

When you respond, please output only one word from the 4 options. Do not include any additional explanations or irrelevant information.

The element is `<image>`. Please predict the given element role:

For training samples, we further include the complete design, canvas size, and element size in the prompt. Such information helps enhance prediction accuracy. For instance, if an element is small in size, it is likely to be an embellishment.

Additional Prompt for Training Samples

The overall design is `<image>`. The canvas width is w_{canvas} px, canvas height is h_{canvas} px. The element is `<image>`. The element width is w_{element} px, element height is h_{element} px. Please also consider the provided canvas and element width/height, as they might be helpful in making a decision. Please predict the given element role:

B. Input-Output Structure in Layered Design Composition

Here we present an example to demonstrate the model input and output in layered design composition. The `<image>` token is a placeholder which is replaced by image embeddings during data loading, as described in Section 3.4. For text elements, we directly put their content in the model input (e.g., *Spring Clean, Best hacks*). Since all intermediate designs can be rendered in advance, the training process is parallelized across the five design layers, with the loss function only calculated in the **ASSISTANT** parts. This ensures the training efficiency of LaDeCo.

C. More Qualitative Results

We show more qualitative results and the layered design composition process in Figure 10 and Figure 11, respectively. The results indicate that LaDeCo is proficient in creating high-quality and visually appealing designs with the novel layered design composition technique.

D. Interesting Applications in Graphic Design

LaDeCo enables interesting applications in graphic design, including design variation, resolution adjustment, and design decoration. We show more results in Figure 12, Figure 13, and Figure 14, respectively.

E. More Qualitative Comparison

Figure 15, 16, 17 show more qualitative comparison of the design comparison, content-aware layout generation, and typography generation tasks, respectively. From these results, we see that LaDeCo significantly outperforms the baseline methods in design composition, and it even surpasses the specialized models in two design subtasks.

A Data Example

HUMAN: a poster of canvas width 1080px, canvas height 1920px. Please predict step by step according to the semantics of the elements. After each prediction, there will be an intermediate rendering result as a reference to better make the next prediction. Now predict the **background** elements: element 0: <image>

ASSISTANT:

```
{
  "index": 0,
  "left": 3,
  "top": -5,
  "width": 1101,
  "height": 460
}
```

HUMAN: current canvas state: <image>. Now predict the **underlay** elements: null

ASSISTANT:

```
{}
```

HUMAN: current canvas state: <image>. Now predict the **logo/image** elements: element 1: <image>

ASSISTANT:

```
{
  "index": 1,
  "left": -78,
  "top": 378,
  "width": 1228,
  "height": 1842
}
```

HUMAN: current canvas state: <image>. Now predict the **text** elements: element 2: *Spring Clean*, element 3: *Best hacks*

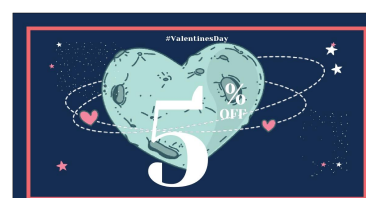
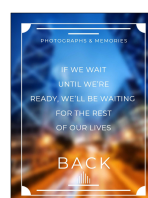
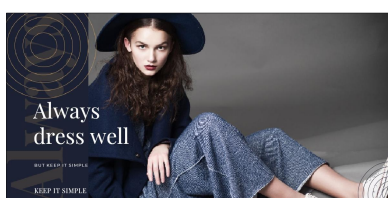
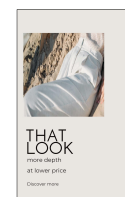
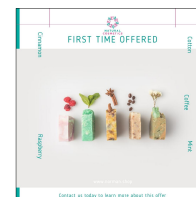
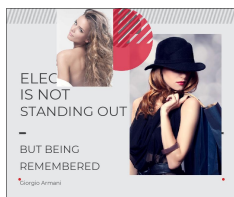
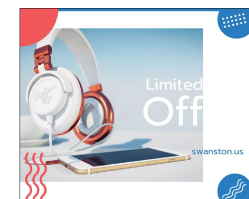
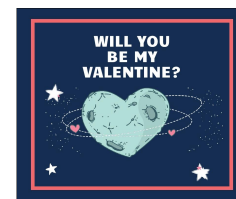
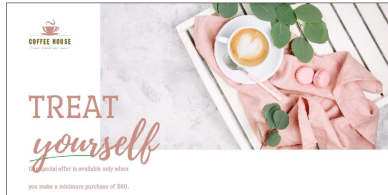
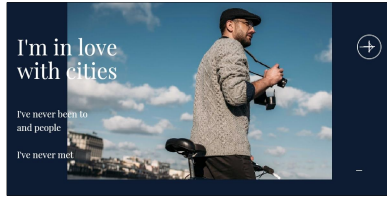
ASSISTANT:

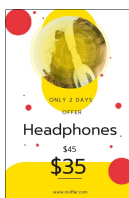
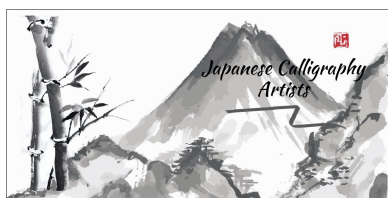
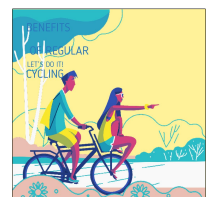
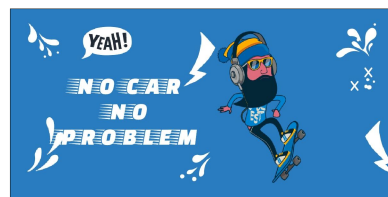
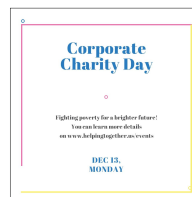
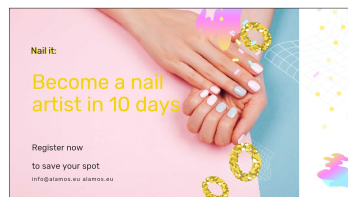
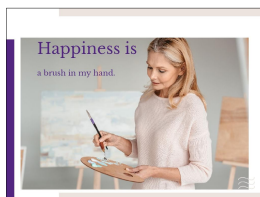
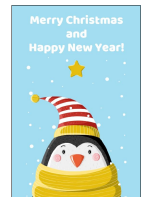
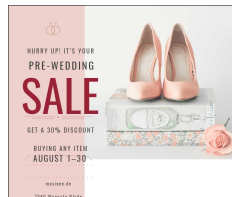
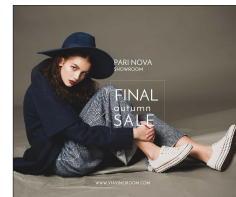
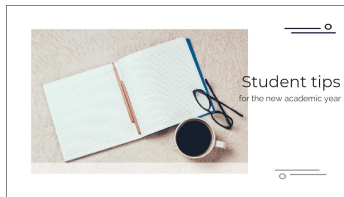
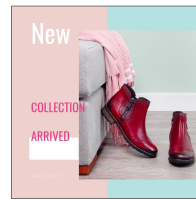
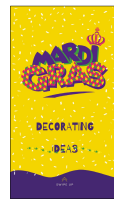
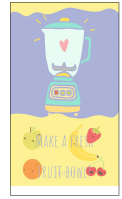
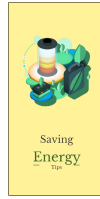
```
{
  "index": 2,
  "left": 98,
  "top": 375,
  "width": 874,
  "height": 125,
  "angle": 0,
  "font": "Raleway",
  "font_size": 125,
  "color": [29, 29, 27],
  "text_align": "center",
  "capitalize": "false",
  "letter_spacing": 0.0,
  "line_height": 1.0,
}
{
  "index": 3,
  "left": 272,
  "top": 547,
  "width": 537,
  "height": 68,
  "angle": 0,
  "font": "Raleway",
  "font_size": 68,
  "color": [0, 0, 0],
  "text_align": "center",
  "capitalize": "false",
  "letter_spacing": 0.0,
  "line_height": 1.0,
}
```

HUMAN: current canvas state: <image>. Now predict the **embellishment** elements: null

ASSISTANT:

```
{}
```





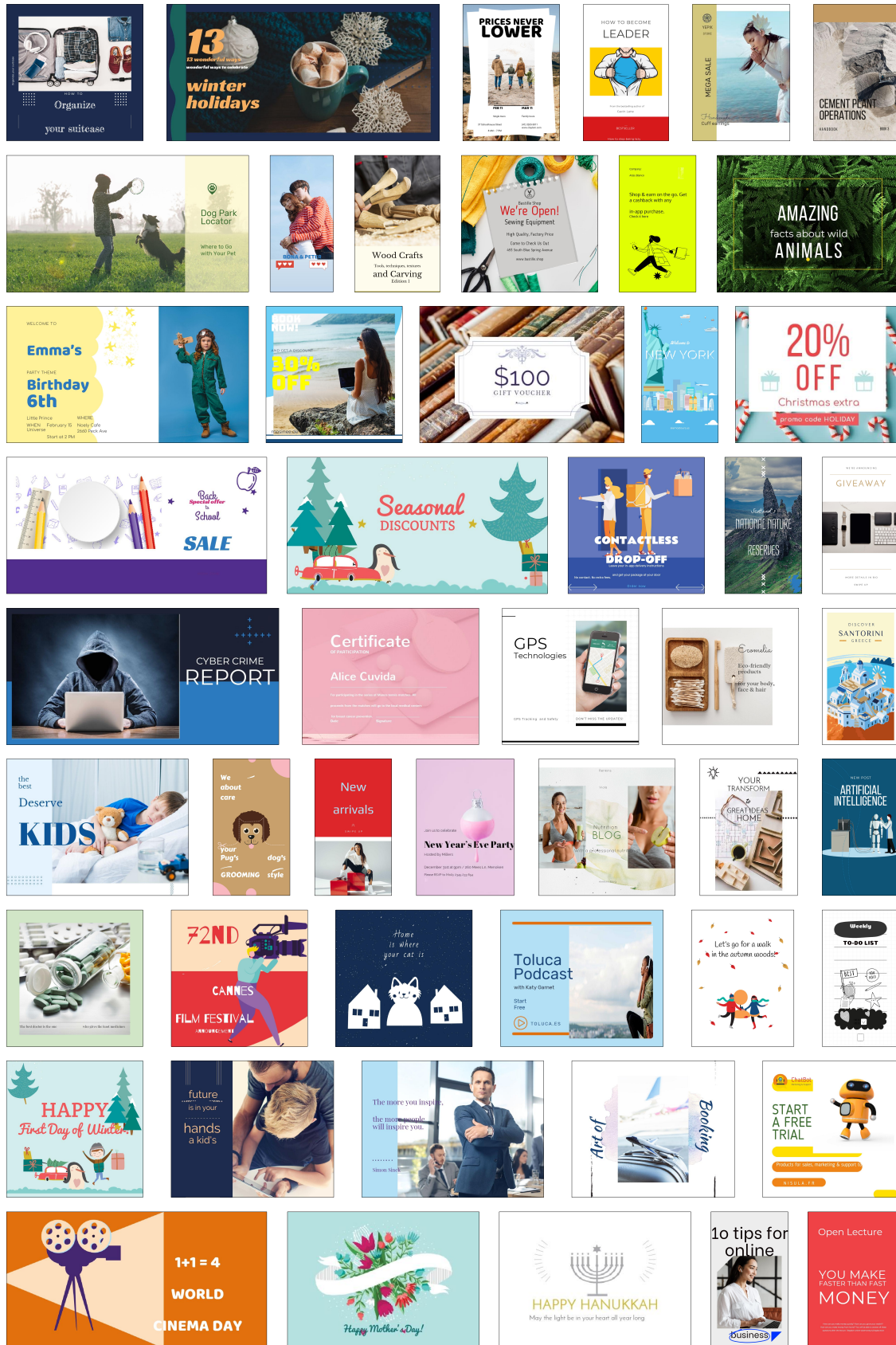
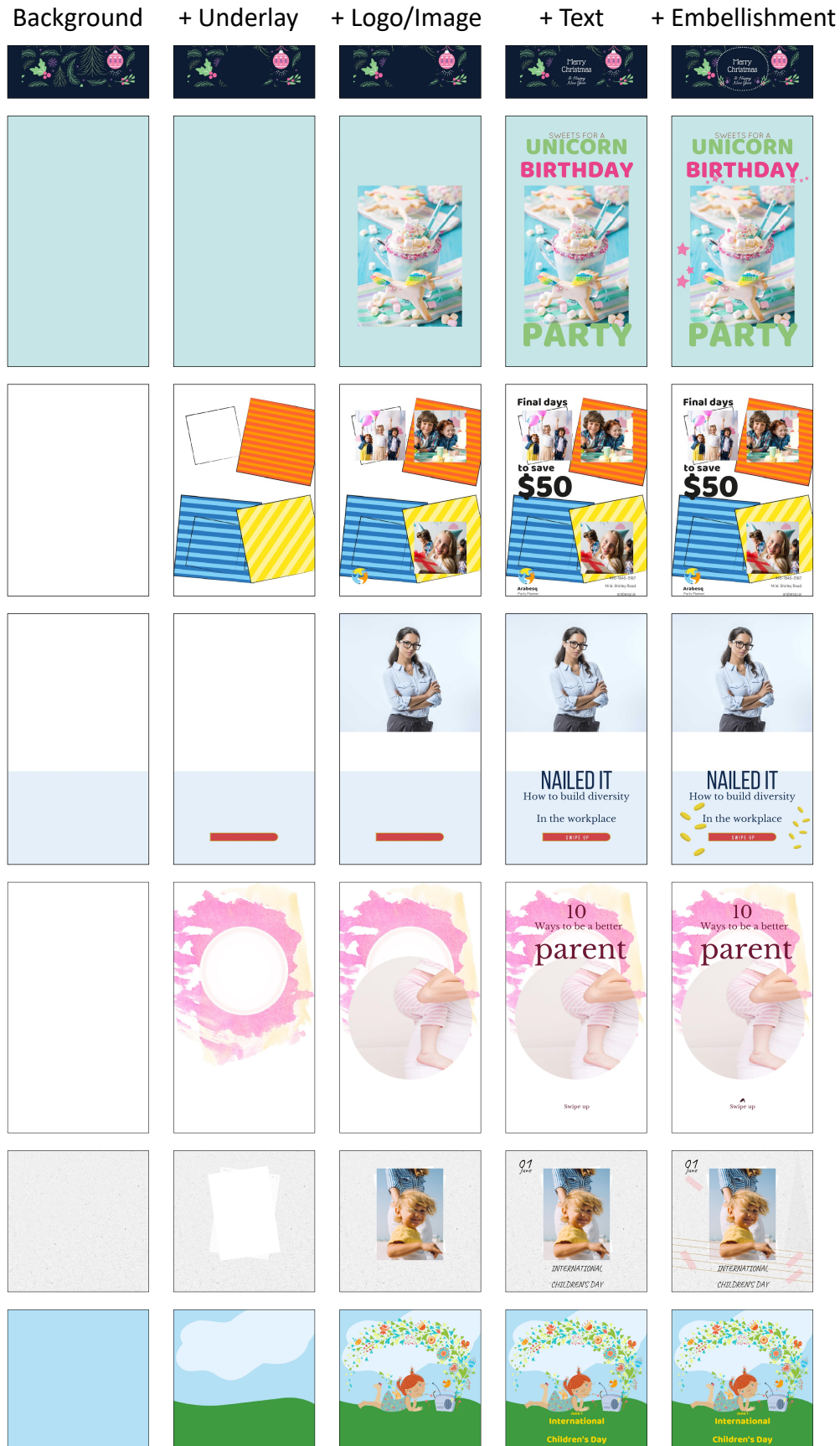


Figure 10. A gallery of graphic designs created by LaDeCo.



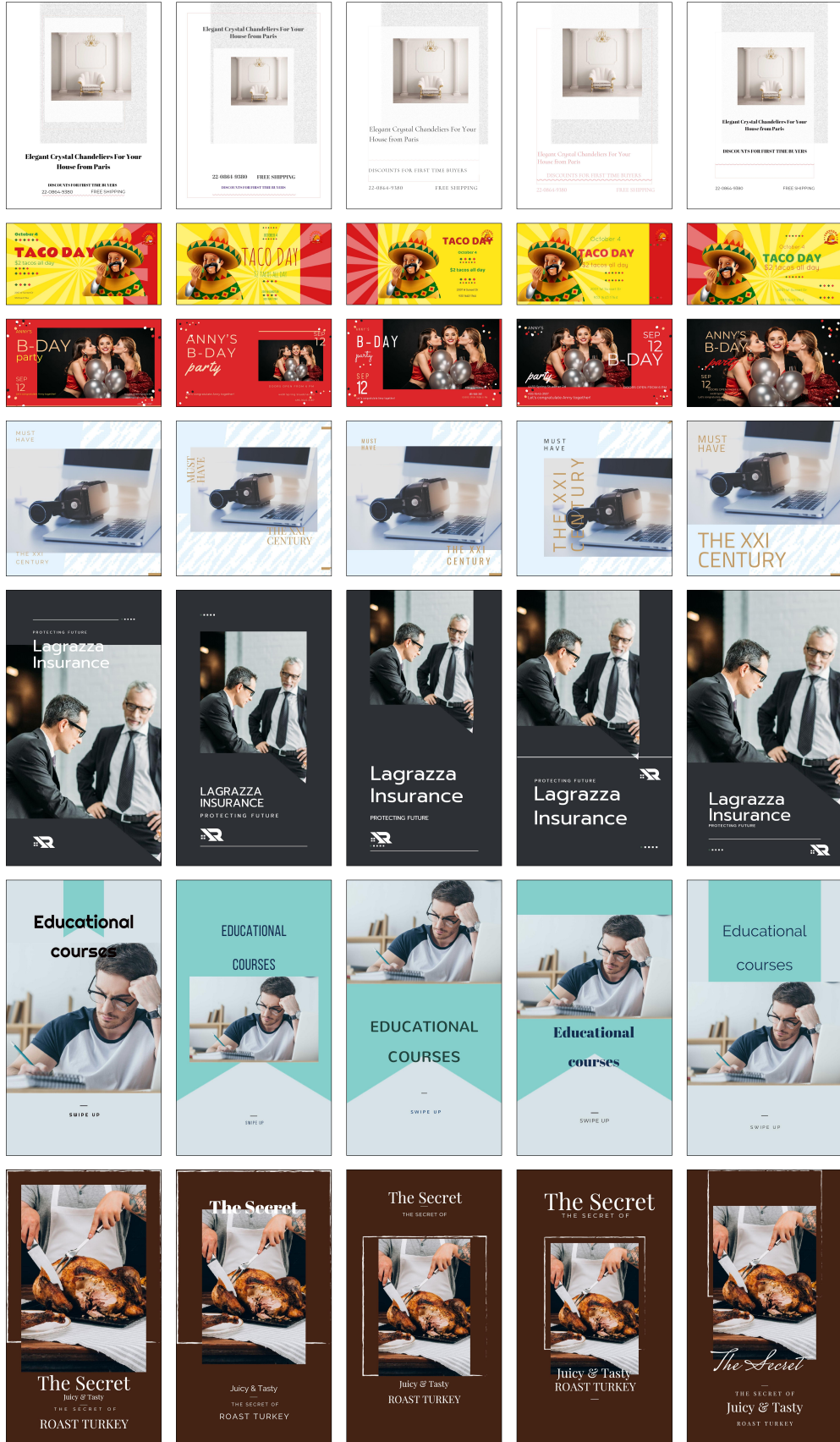
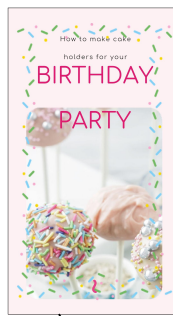
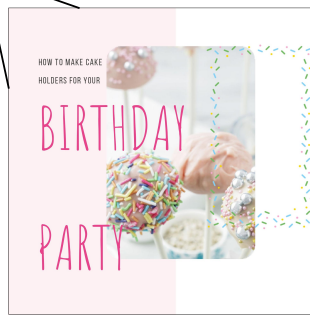


Figure 12. More results to demonstrate that LaDeCo can create diverse designs with the same input.

1920px



1080px

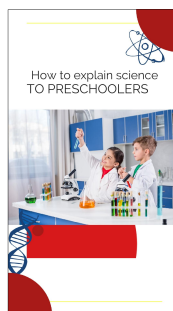
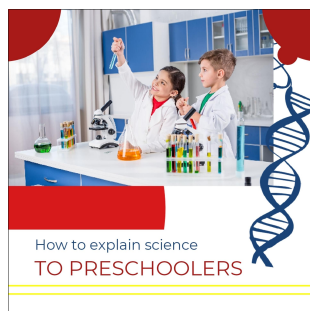
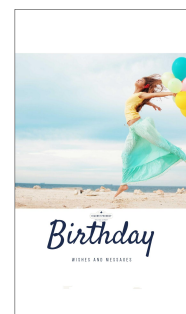
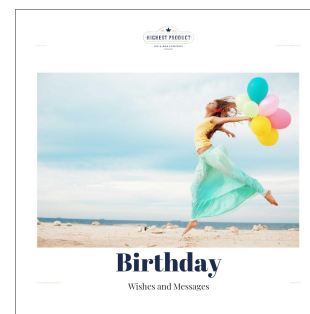
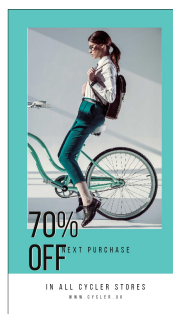
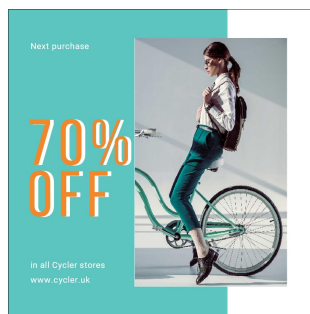
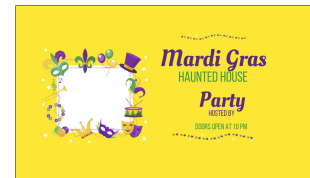


Figure 13. More results to demonstrate that LaDeCo is able to generate graphic designs with different aspect ratios.

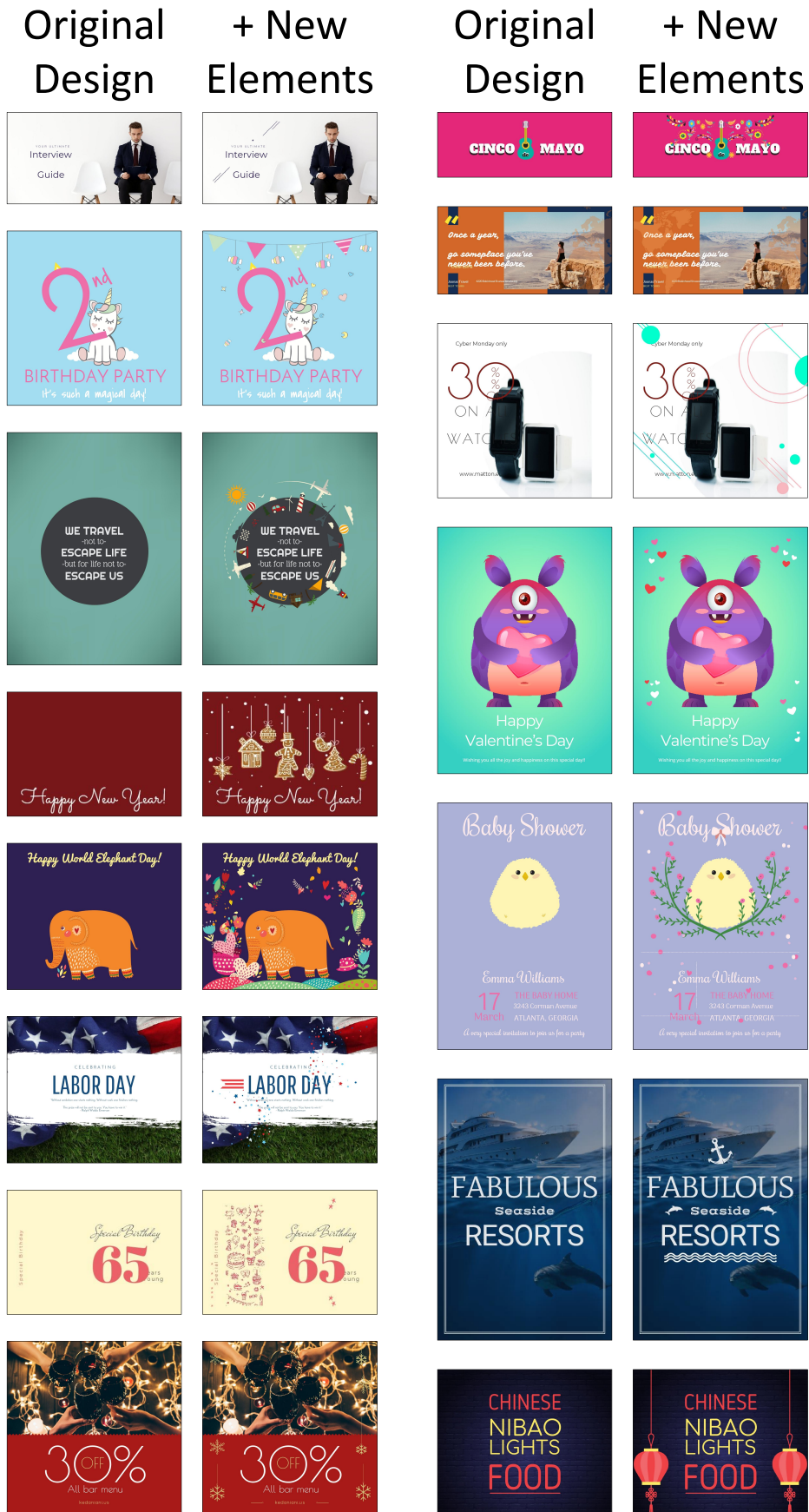


Figure 14. More results to demonstrate that LaDeCo can add new elements to an existing design in a plausible way.

FlexDM

GPT-4o

Ours

GT

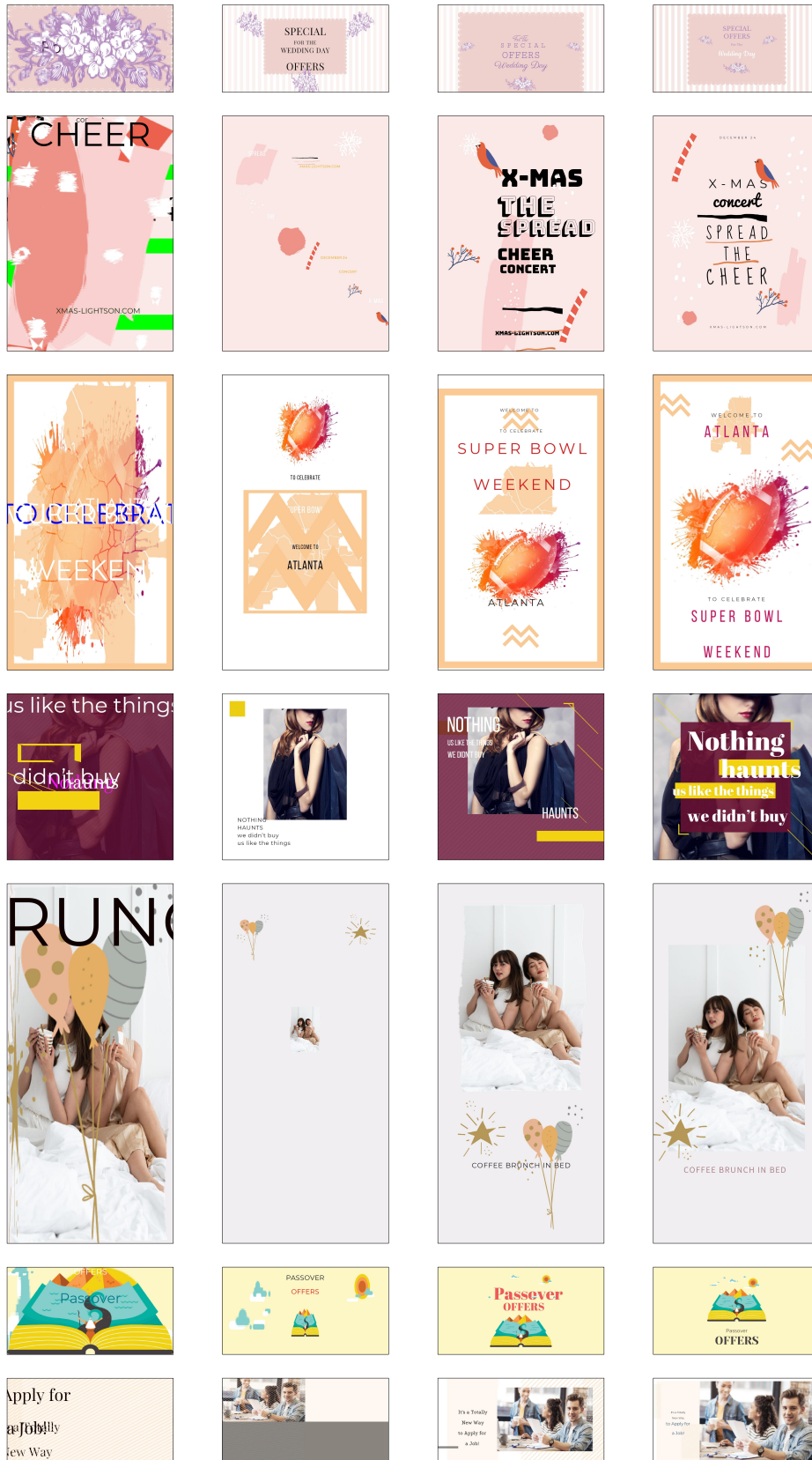


Figure 15. More qualitative comparison to demonstrate the superiority of LaDeCo in design composition.

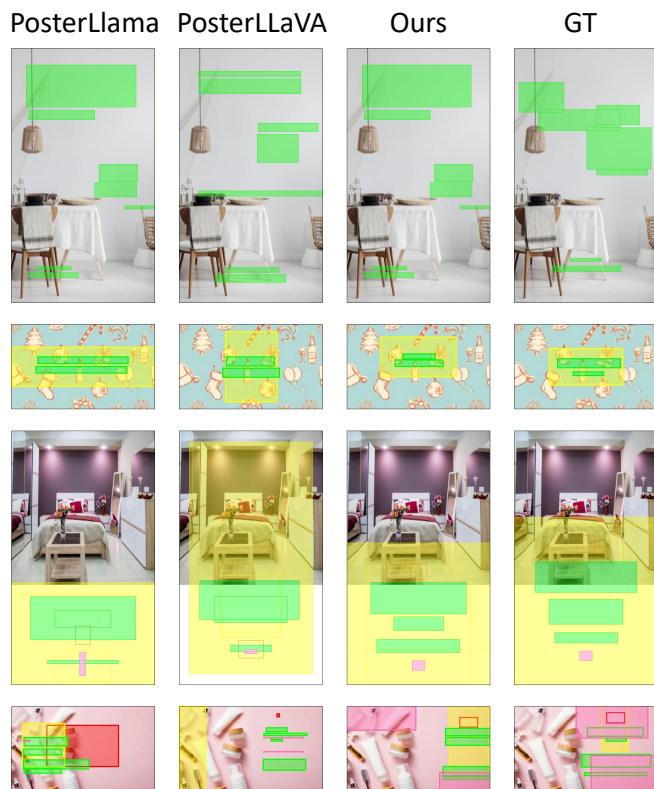


Figure 16. More qualitative comparison to demonstrate the superiority of LaDeCo in content-aware layout generation.

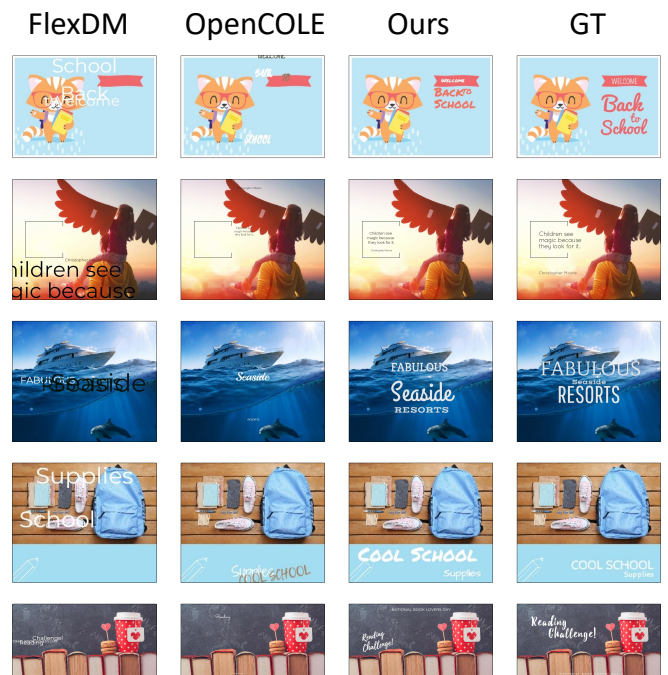


Figure 17. More qualitative comparison to demonstrate the superiority of LaDeCo in typography generation.