Desigen: A Pipeline for Controllable Design Template Generation (Supplementary Material)

Haohan Weng¹^{*}, Danqing Huang², Yu Qiao³, Zheng Hu³^{*}, Chin-Yew Lin², Tong Zhang¹, C. L. Philip Chen¹ ¹South China University of Technology, ²Microsoft, ³Central South University



Figure 1. Samples of design from the proposed dataset.

1. Design Dataset

To the best of our knowledge, there is no current available dataset including content descriptions, background images, and layout metadata. Therefore, we crawl home pages from online shopping websites and extract the banner sections, which are mainly served for advertisement purposes with well-designed backgrounds and layouts, leading to a new dataset of advertising banners. The collected banners contain background images, layout element information (i.e., type and position), and text descriptions. Several sample designs are shown in Figure 1. In Figure 2, we demonstrate the word cloud of the top 200 words of corresponding text descriptions, showing that descriptions are highly relevant to the graphic designs.

2. More Qualitative Results

In this section, we show the additional results for the proposed model. Figure 4 shows the additional generated banners via the proposed layout generator, where the blue boxes represent text elements, the red boxes represent button elements and the green boxes represent image elements. It shows that our layout generator performs well on both real



Figure 2. Word cloud of the top 200 words of text descriptions on the proposed design dataset.

backgrounds and generated backgrounds. In Figure 3, the proposed model can synthesize varying background images given the same text prompts. In Figure 7, additional results are shown for slide deck generation with the same prompts and different attention reduction masks. Figure 6 shows additional attention visualization of background-aware layout generation. The layout generator highly attends to the salient objects of the background images and creates layouts on the region with low attention scores.



Figure 3. The proposed model can synthesize varying background images given the same text prompts. The prompt of the first row: green plants with white background; the second row: Cute little hamster with flower; the third row: Squirrel in the snow.



(a) Designs based on real background

(b) Designs based on generated background

Figure 4. More synthesized layouts via the proposed layout generator, where the blue boxes represent text elements, red boxes represent button elements and green boxes represent image elements.



Figure 5. More qualitative results of complicated layouts (more than four elements), where the blue boxes represent text elements, red boxes represent button elements, and green boxes represent image elements.



Figure 6. Additional attention visualization of background-aware layout generation. The layout generator highly attends to the salient objects of the background images and creates layouts on the region with low attention scores.



Figure 7. Additional results for slide deck generation with the same prompts and different attention reduction masks (Prompt: Rose for Valentine's Day).