Visual Layout Composer: Image-Vector Dual Diffusion Model for Design Layout Generation

Supplementary Material

In the supplementary section, we offer extra qualitative findings, expanded information, and further comparisons between the Poster and Crello datasets. We also discuss instances where our method does not work well, and introduce an additional perceptual metric based on a specialized visual classifier.

6. Additional Qualitative Results

In Figure 13 and Figure 14, we present a further qualitative comparison between our dual-domain model and the vector-only model. Although the vector model is capable of placing bounding boxes for various elements in appropriate positions, its insufficient visual reasoning results in outputs of lower design quality compared to those produced by our dual-domain model.

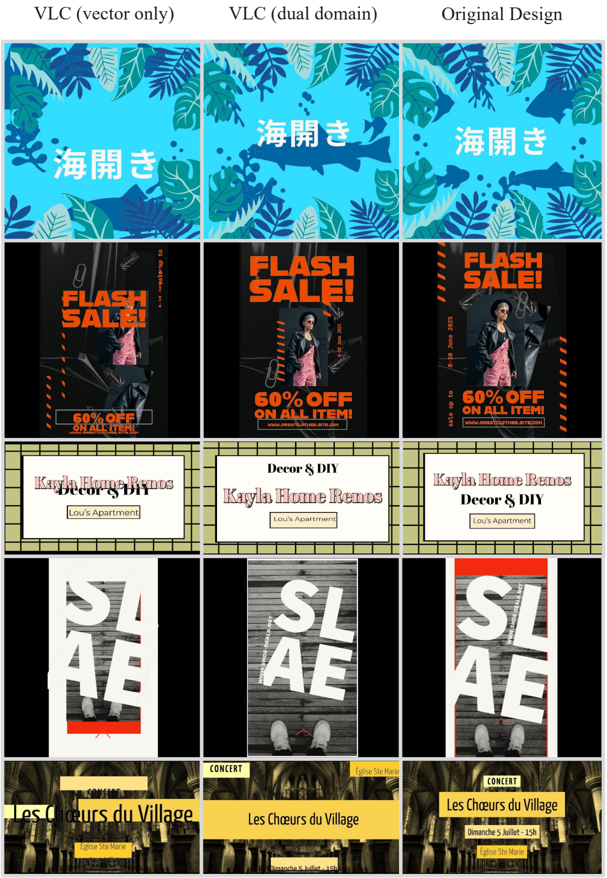

Figure 13. Additional qualitative comparison between our method and baselines on Poster dataset.

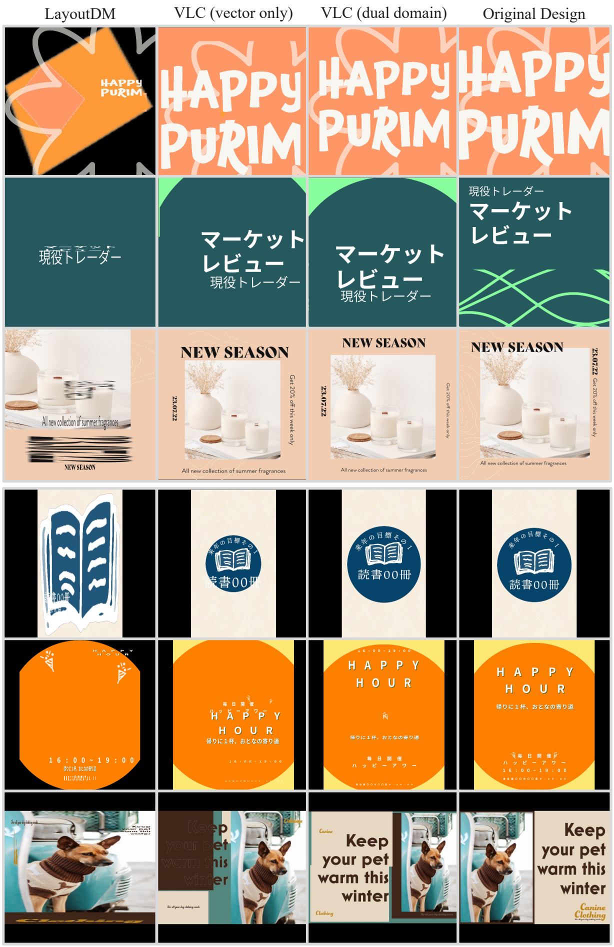

Figure 14. Additional qualitative comparison between our method and baselines on Crello dataset.

7. Failure Cases

Figure 15 illustrates some limitations of our method. A significant issue with our model arises when dealing with repetitive, abstract elements. This scenario leads to the attention layers concurrently focusing on all similar elements, resulting in the omission of certain elements in the image output. Additionally, the attention scores in the vector domain become convoluted, as they encapsulate all these similar elements in each attention map. Although we've mitigated this issue with Attention Localization Loss, it still presents challenges, particularly in layout designs like menus. Furthermore, layout designs that incorporate elements with extreme aspect ratios, such as elongated lines or

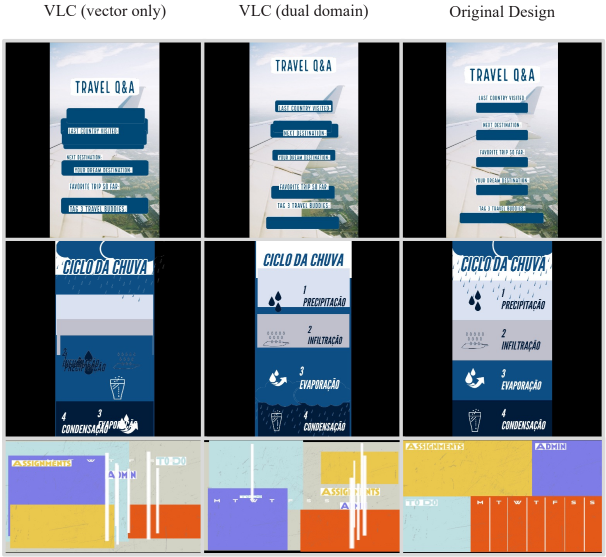

Figure 15. Most failure cases of our method include elements with extreme aspect ratios or those featuring repetitive abstract elements, like menus.

very small components, result in disproportionately small attention maps. These maps may be effectively excluded from the 8×8 attention scores received by the vector domain from the image domain's middle block. Consequently, this disparity between the domains leads to inconsistencies and potentially corrupts the diffusion process. We will explore how to fix these issues with more regularization terms and higher resolution maps of the attention layers.

8. Dataset Comparison

Figure 17 shows additional examples from both Poster and Crello datasets. Although Crello dataset has up to 30 elements per layout, for Poster dataset, we include up to 20 elements based on the complexity of the designed layouts. We show the distribution of number of training samples based on the number of elements in Figure 16. There are 133,781, and 18,768 training samples respectively in Poster and Crello datasets.

9. Classifier Evaluation

The FID score used for evaluating generated layout images is very limited given the fact that it only measures the image feature distribution globally without comparing the results sample by sample. It is also sensitive to image content feature and local patch details, which are irrelevant to the layout quality studied in this paper. Therefore, we introduce an alternative metric by training a classifier to assess each generated layout image individually. This binary classifier is trained to distinguish between "real" and "fake" layout images rendered with ground truth and randomly perturbed layouts, respectively. We compare the average classifier scores and rankings for the testing results generated by different methods. The score is calculated as the predicted probability of real class, and the ranking is calculated as $1/\log_2(1+r)$ where r is the ranking index among the compared methods. The ranking metric is similar as the ranking weight used in normalized discounted cumulative gain (NDCG).

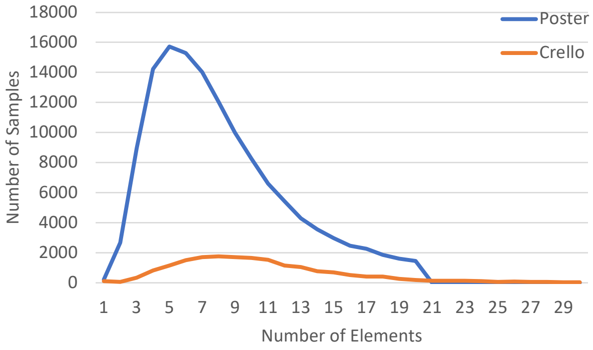

Figure 16. Distribution of the number of training samples with respect to the number of elements in Poster and Crello datasets.

Table 3 shows the superior performance of our dual-domain model as evidenced by its higher scores in comparison to the vector-only approach and LayoutDM. Our classifier is trained on Poster dataset which includes higher quality posters and elements, therefore leading to lower overall score for Crello dataset. The ranking metric is not affected by the classifier bias and shows similar values for the two datasets.

Table 3. Visual design quality comparison of different methods based on classifier score and ranking.

Dataset	Method	Score	Ranking
Crello	LayoutDM [17]	0.462	0.549
	Vector-only (Ours)	0.632	0.644
	Dual-domain (Ours)	0.645	0.662
	Original Designs	0.667	0.706
Poster	LayoutDM [17]	0.612	0.565
	Vector-only (Ours)	0.753	0.635
	Dual-domain (Ours)	0.770	0.668
	Original Designs	0.789	0.693

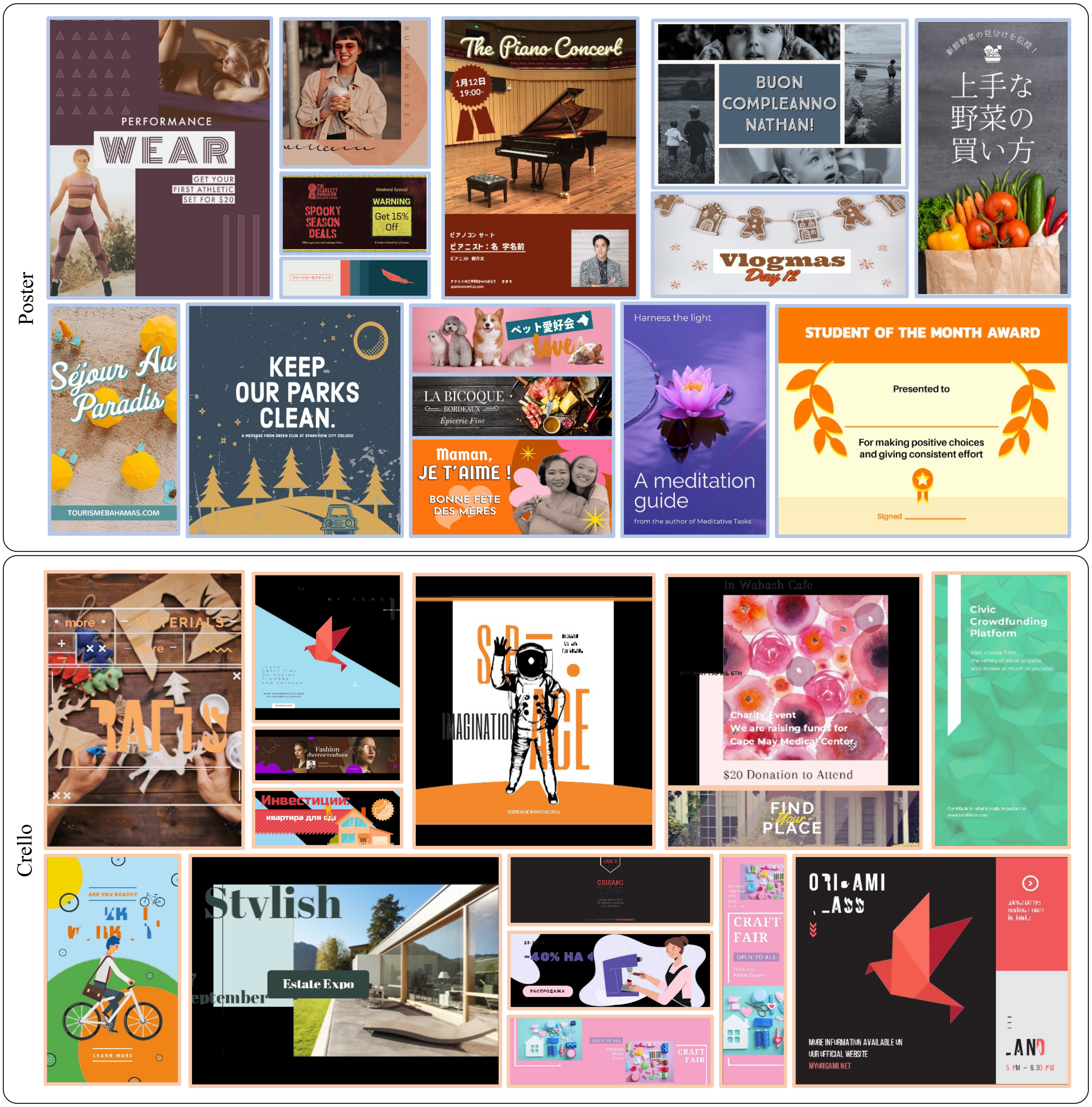

Figure 17. The Poster dataset at the top features original layout designs of higher quality with a diverse set of elements, while the Crello dataset at the bottom includes layout templates that exhibit repetitive elements between the designs.