Design-o-meter: Towards Evaluating and Refining Graphic Designs (Supplementary Material)

Sahil Goyal*[†], Abhinav Mahajan*[‡], Swasti Mishra, Prateksha Udhayanan, Tripti Shukla,
K J Joseph, Balaji Vasan Srinivasan
[†]IIT Roorkee [‡]IIIT Bangalore Adobe Research
sahil_g@ma.iitr.ac.in, abhinav.mahajan@iiitb.ac.in, {josephkj, balsrini}@adobe.com

A. Additional Ablation Experiments

A.1. Alternate Margins for Eqn. 3

We experiment with Hard margin (H-Margin), transformation-based margin (TB-Margin), and adaptive margin (Ada-Margin). For H-Margin, we try low (0.2), medium (0.5), and high (1.0) values. A low value of 0.2 achieves the highest RAcc of **94.97**. For TB-Margin, we assign different margins to different transformations, a low value (0.2) for transformations adding noise with a low std deviation, and similarly for medium (0.4) and high (0.6) noises. We define Ada-Margin as:

$$\text{Ada-Margin} = \max\left(\max_{(g_i,b_i)\in\text{batch}} \lambda\left\|\mathcal{F}(g_i) - \mathcal{F}(b_i)\right\|_2, 0.1\right)$$

where λ is a scaling factor, we choose $\lambda = 0.05$.

A.2. Alternate Similarity Loss Formulation

We experiment with different similarity losses (exponential, binomial deviance, and square) for Eqn. 4.

$$L_{\text{sim}}^{\text{exp}} = e^{P_{\text{sim}}(\mathcal{S}(\boldsymbol{D}_{meta}^{good}),\ \mathcal{S}(\boldsymbol{D}_{meta}^{bad}))} \tag{1}$$

$$L_{\text{sim}}^{\text{dev}} = \ln(e^{2*P_{\text{sim}}(\mathcal{S}(\boldsymbol{D}_{meta}^{good}),\ \mathcal{S}(\boldsymbol{D}_{meta}^{bad})} + 1) \tag{2}$$

$$L_{\text{sim}}^{\text{sq}} = (P_{\text{sim}}(\mathcal{S}(\boldsymbol{D}_{meta}^{good}),\ \mathcal{S}(\boldsymbol{D}_{meta}^{bad})) + 1)^2 \tag{3}$$

P_{sim} is an embedding similarity computed as the dot product between the tanh-activations of the "good" and "bad" design pairs as follows:

$$P_{sim} = \frac{\mathcal{F}(\boldsymbol{D}_{meta}^{good}).\mathcal{F}(\boldsymbol{D}_{meta}^{bad})}{\max(\left\|\mathcal{F}(\boldsymbol{D}_{meta}^{good})\right\|_2 . \left\|\mathcal{F}(\boldsymbol{D}_{meta}^{bad})\right\|_2, \epsilon)}; \epsilon > 0 \tag{4}$$

We get similar RAcc using all the similarity losses with minor differences. We choose $L_{\text{sim}}^{\text{dev}}$ because of the validation loss decreases most in this case.

A.3. Use of Classifier Guidance in Scorer

We try to guide the scorer model with a binary classification head on top of the siamese model. We additionally introduce a binary cross-entropy loss to differentiate good and bad designs. This setting increases the model size and training time, but doesn't significantly help the scorer model in ranking the designs better.

B. Failure Cases

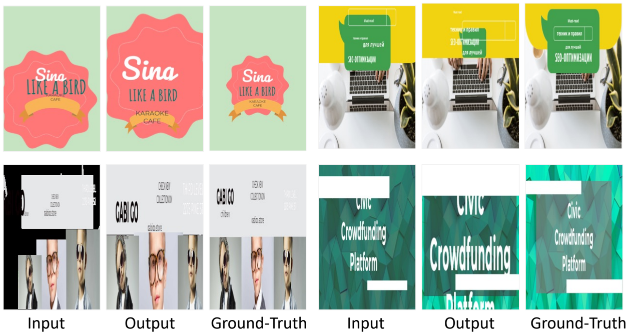

Figure 1. Examples where Design-o-meter fails to achieve the optimal refinement.

We showcase failure cases of Design-o-meter in Fig. 1, ranging from minor to significant failures. Despite being an excellent scorer and refiner, at-times the signals from the input are weak to correctly guide the layout and scale transformations.

C. Perturbations used in Dataset Creation

In Fig. 4, we show a visualization of the perturbations that we do to the input design to create the dataset to train the *scorer* module, as explained in Sec. 3.1.

D. Additional Results

We include more qualitative results on the scores and refined outputs in Figs. 2 and 3.

*Equal contribution. Work done during internship at Adobe Research.

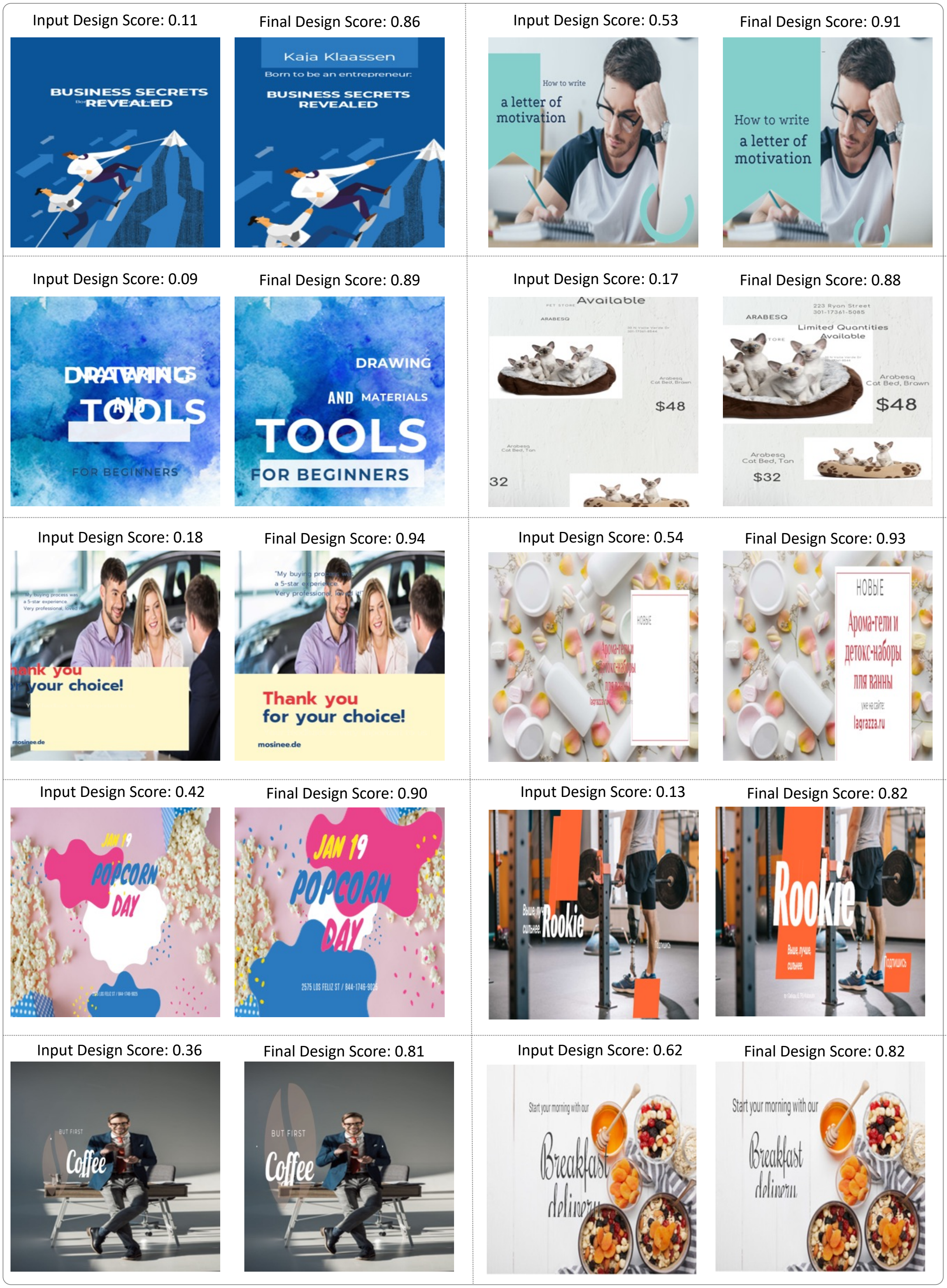

Figure 2. We add more qualitative results here. The input design and its corresponding refined output along with the scores are shown.

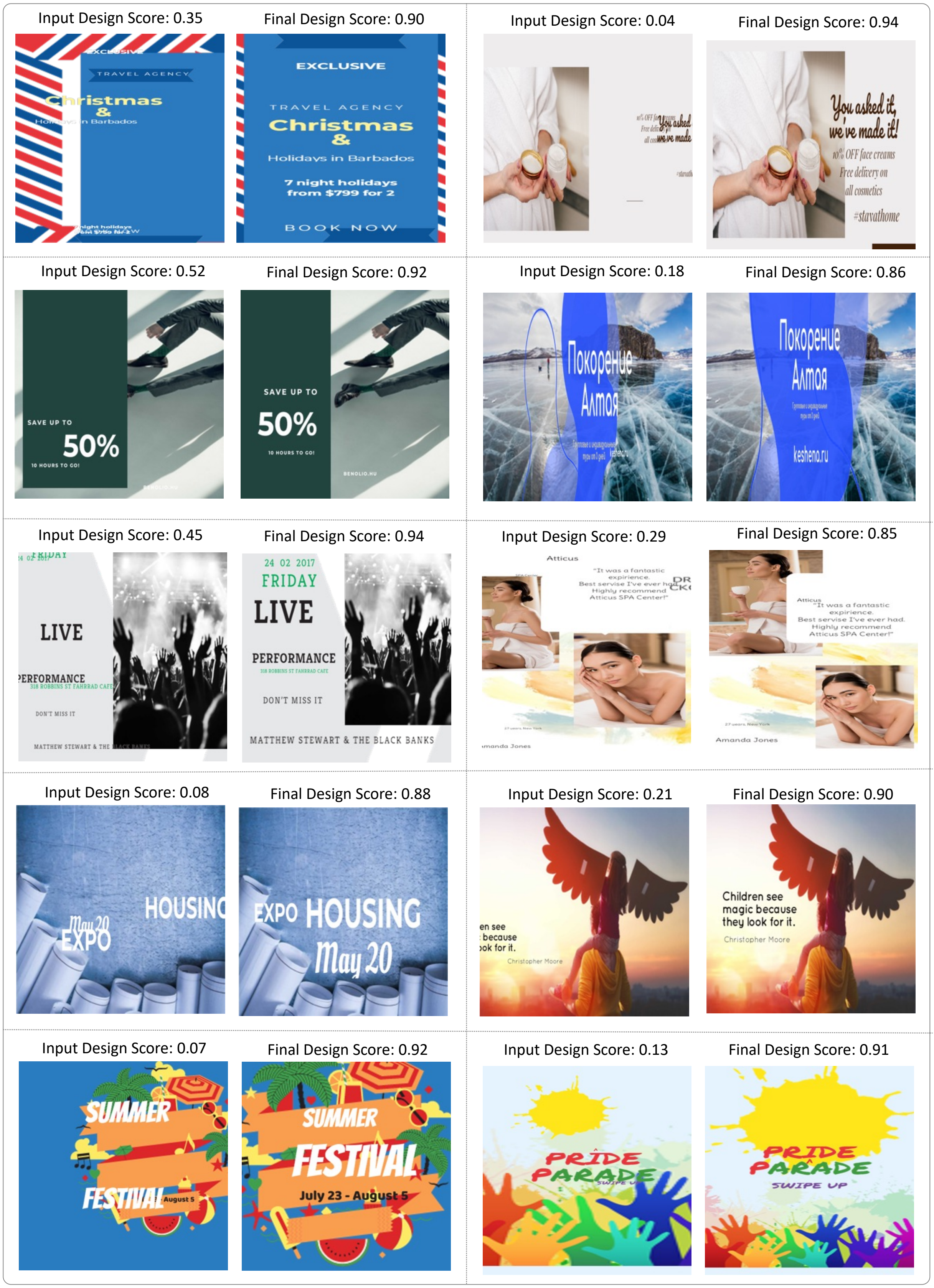

Figure 3. We add more qualitative results here. The input design and its corresponding refined output along with the scores are shown.

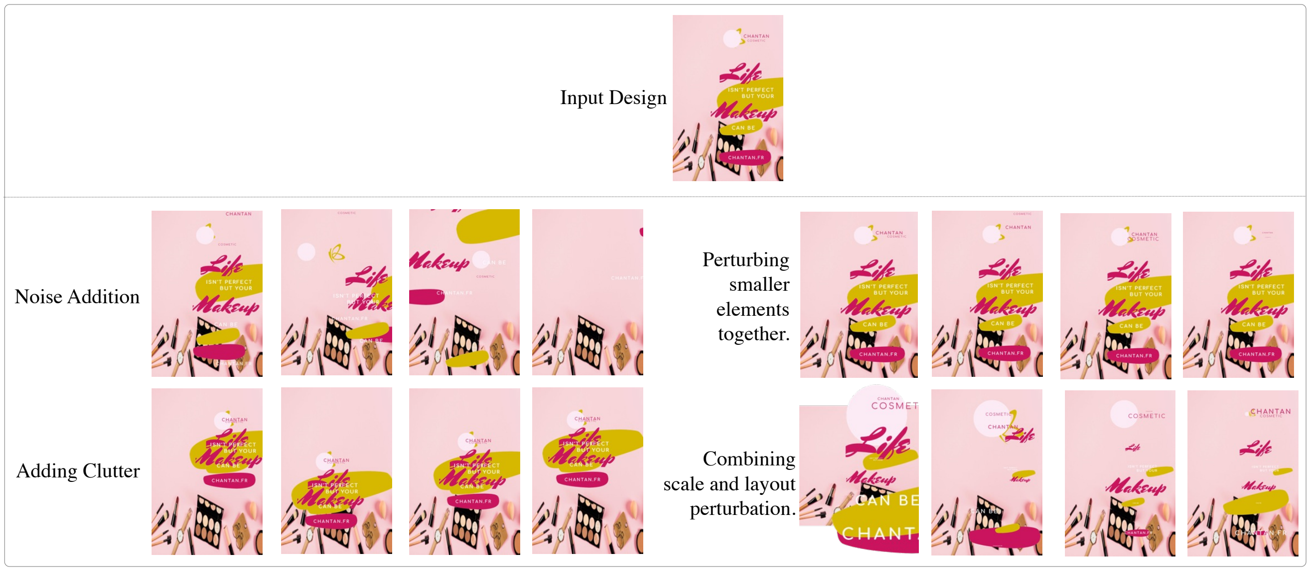
Input Design
Noise Addition
Perturbing smaller elements together.
Adding Clutter
Combining scale and layout perturbation.

Figure 4. We show the perturbations that we apply to input design, for creating {good-design, bad-design} pairs to train our *scorer* model.