Supplementary Material: Hierarchical Layout-Aware Graph Convolutional Network for Unified Aesthetics Assessment

Dongyu She[1], Yu-Kun Lai[2], Gaoxiong Yi[3], Kun Xu[1]*
[1]Tsinghua University [2]Cardiff University [3]Tencent

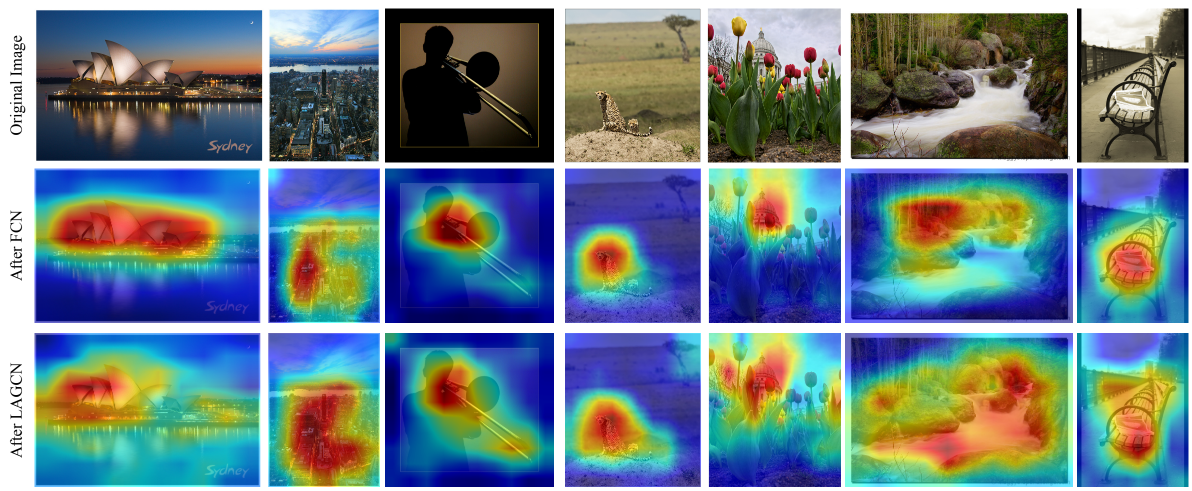

Figure 1. Qualitative results of Class Activation Maps after FCN (second row) and after the first LAGCN module (third row), respectively.

In the supplementary material, we present more visual results on the benchmark AVA dataset. Note that all images are shown with the original aspect ratios preserved.

1. Class Activation Maps

We first show more examples of the Class Activation Maps (CAMs) using features after FCN and after the first LAGCN module from our fine-tuned HLA-GCN, respectively. As observed in Fig. 1, the features after graph convolution have larger regions of interests (highlighted in red) that cover highly correlated scenes.

2. Mean Aesthetics Score Prediction

Fig. 2 and Fig. 3 show images predicted with low assessment errors and large prediction errors for the aesthetics score prediction task, respectively. Generally, we find that images with low errors have high technical quality and show the prominent subject, while images with large prediction errors tend to have uncommon characteristics, *e.g*. abstract style, non-object texture and large depth of field.

3. Aesthetics Distribution Prediction

Fig. 4 shows the predicted aesthetics distributions using our proposed model. As observed, our HLA-GCN can predict aesthetics distributions consistently with the ground-truth distributions. Fig. 5 shows some cases with large assessment errors. Our model performs worse on images whose aesthetics distributions are highly non-Gaussian.

4. Model Comparison

Finally, we compare the prediction results using different models, *i.e*. the NIMA framework and our proposed HLA-GCN. In Fig. 6, both models work well on the generic images (first row), while our proposed model is more effective for learning layout-aware representations (last row).

*Corresponding author, xukun@tsinghua.edu.cn

Figure 2. Mean aesthetics score predictions of the proposed HLA-GCN from the AVA test set, for which the Mean Squared Errors (MSE) are the lowest. Our predicted scores are given below each image while the ground-truth scores are given in brackets.

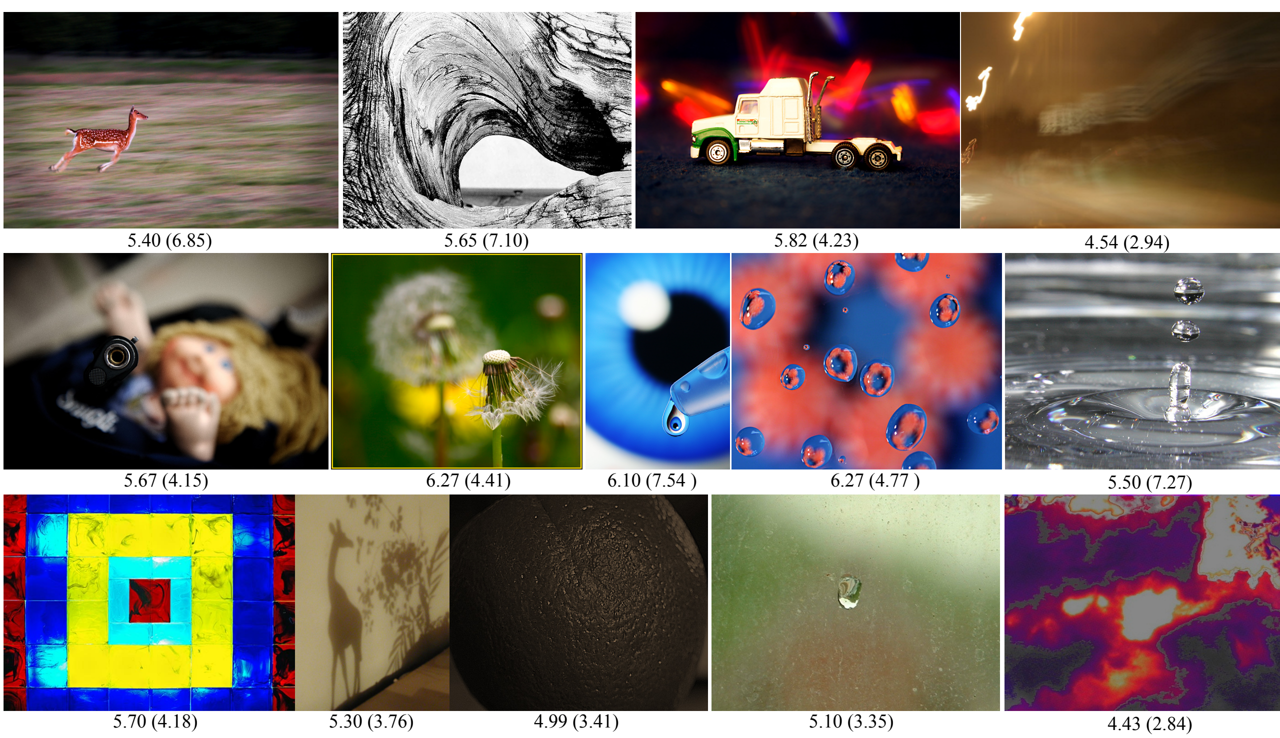

Figure 3. Mean aesthetics score predictions of the proposed HLA-GCN from the AVA test set, for which the MSEs are some of the highest. Our predicted scores are given below each image while the ground-truth scores are given in brackets.

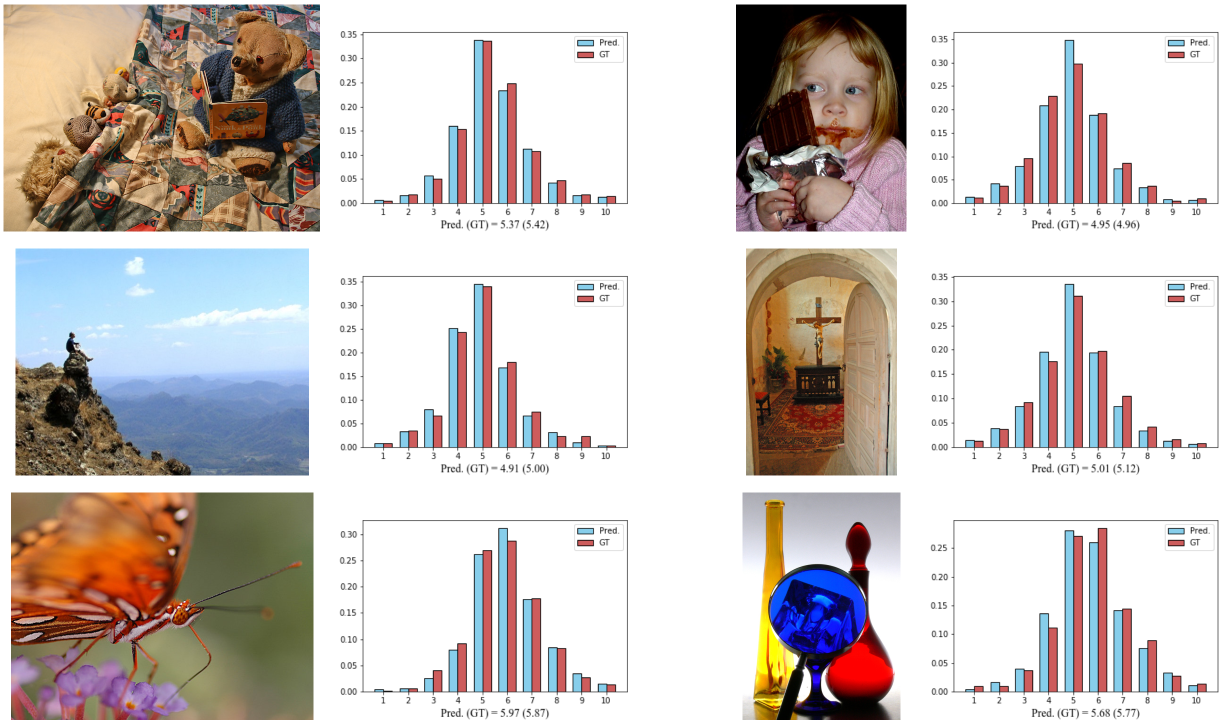

Figure 4. Aesthetics distribution predictions of our model (Pre.) and the groundtruth (GT), for which the squared Earth Mover's Distance (EMD) errors are the lowest. The corresponding mean aesthetics scores are also shown below while the ground-truth are given in brackets.

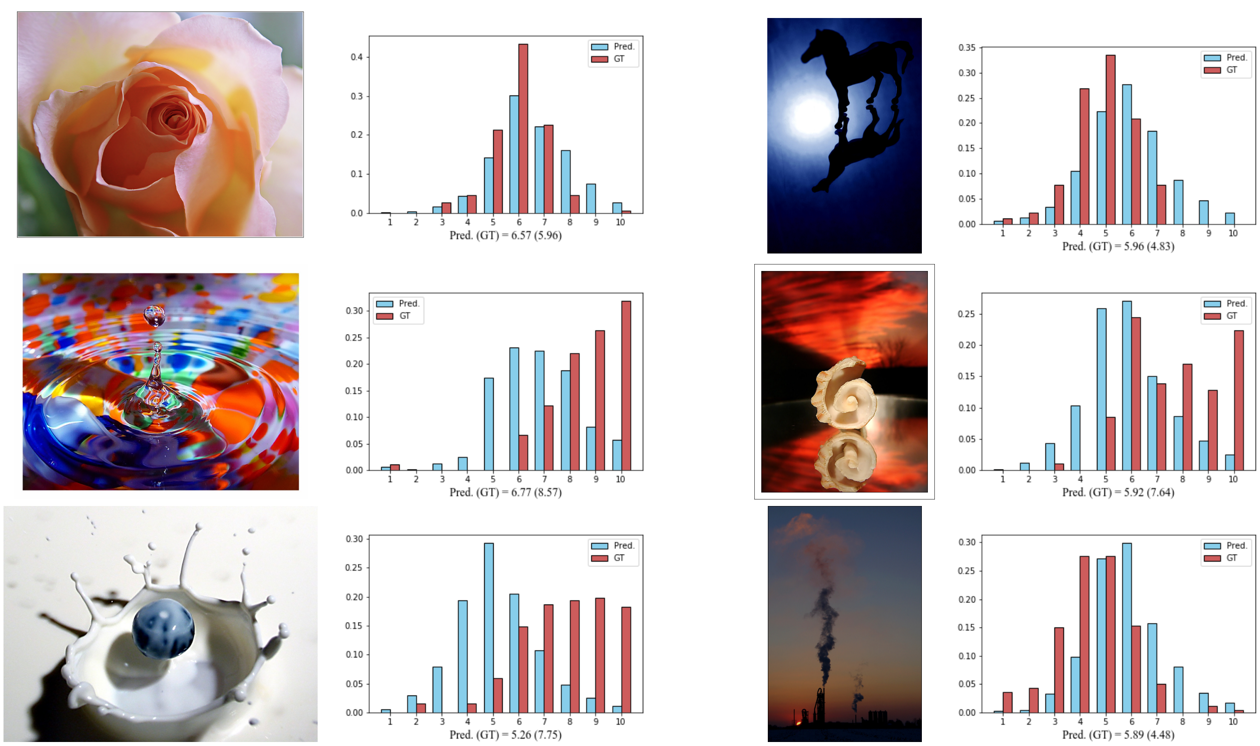

Figure 5. Aesthetics distribution predictions of our model (Pre.) and the groundtruth (GT), for which the EMD errors are some of the highest. The corresponding mean aesthetics scores are also shown below while the ground-truth are given in brackets.

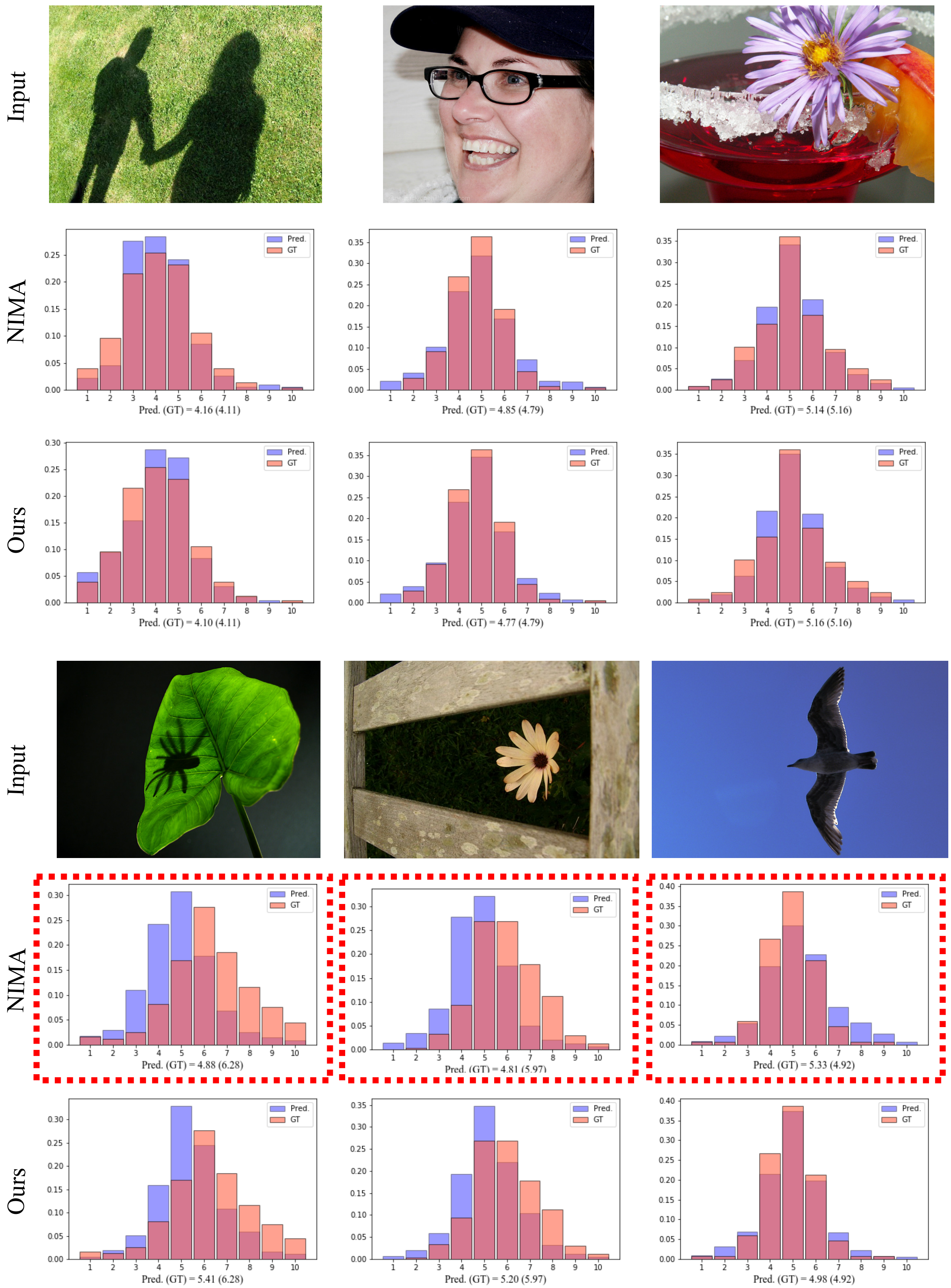

Figure 6. Aesthetics distribution predictions using the NIMA framework and our model. Each sub figure is coupled with plots of the groundtruth and predicted score distributions. The corresponding mean aesthetics scores are also shown below while the ground-truth scores are given in brackets. Failure cases for the binary aesthetics quality classification task are denoted in the boxes with red dash lines.