Supplementary Material
Composing Photos Like a Photographer

Chaoyi Hong[†] Shuaiyuan Du[†] Ke Xian[†] Hao Lu[†] Zhiguo Cao[†,*] Weicai Zhong[‡]
[†]Key Laboratory of Image Processing and Intelligent Control, Ministry of Education
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology
[‡] Huawei CBG Consumer Cloud Service Search Product & Big Data Platform Department
{cyhong, zgcao}@hust.edu.cn

This document includes the following contents:

1. Technical architecture of CACNet;
2. Classification accuracy of the composition branch;
3. Sensitivity experiments of the balancing factor λ;
4. Qualitative comparisons with other methods and visual examples of the interpretable image cropping;
5. Details of the user study.

1. Technical Architecture

Here we present the technical architecture of CACNet in Fig. 1, which comprises three parts: a backbone, a composition branch, and a cropping branch. The backbone follows [8, 9, 6], which adopts all convolution blocks excluding the last max pool layer of VGG-16 [7] and produces 512-dimensional (channels) features of $\frac{1}{16}$ input resolution (512-d, 16-r). The output of the backbone flows into both the composition and cropping branches. The composition branch includes a decoder, a global average pooling (GAP) layer, and a fully-connected (FC) layer. The decoder first applies two 256-d 3×3 conv layers, each of which followed by a BatchNorm layer and a ReLU function. Then it upsamples $\times2$ the feature map and element-wise add the feature map with the output of VGG-16 $pool3$. After are the following operations, a 128-d 1×1 conv layer, upsampling $\times2$, element-wise adding with the output of $pool2$, and a 128-d 1×1 conv layer. Finally the 128-d, 4-r feature map is fed to a GAP and a FC layer to produce the confidence scores across different rules of composition. The cropping branch applies three groups of layers, each consisting of one 256-d 3×3 conv, one BatchNorm, and one ReLU layer. Finally, a $((\frac{16}{K})^2 * 4)$-d 3×3 conv layer is applied to predict the offsets between the anchors and the ground truth coordinates, where K is the stride of anchors.

[*]Corresponding author.

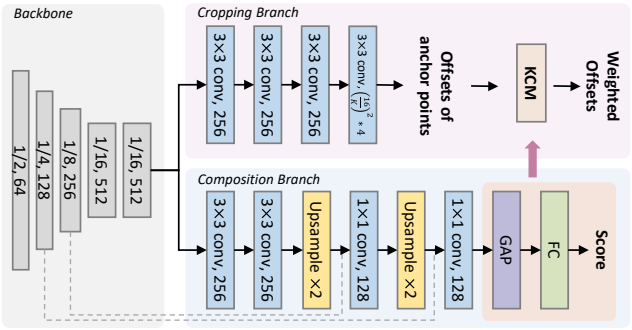

Figure 1. Technical architecture of CACNet.

Table 1. Classification accuracy in overall and across 9 composition rules.

Rules	Overall	Cen.	Hor.	Dia.	Tri.	RoT	Sym.	Cur.	Ver.	Pat.
Acc.(%)	**87.8**	91.4	90.1	71.5	87.2	94.4	79.8	60.9	86.4	93.7

2. Accuracy of Composition Branch

The classification accuracy of the composition branch is crucial to the calculation of KCM, and further the cropping results. Here we show the classification accuracy on KU-PCP dataset [4] in Table 1. Considering that one image may be tagged with more than one class, it is deemed as correctly categorized if predicted as one of the ground-truth composition classes. Overall the classification accuracy is 87.8%. The high accuracy justifies the effectiveness of the composition branch to recognize and distinguish different composition patterns. In specific, the high accuracy of RoT, Cen., Hor., and Pat. can boil down to distinctive and simple leading elements whereas the low accuracy of Sym., Dia., especially Cur., may lie in the complicated ones. Generally, the classification accuracy is sufficient to help produce appropriate KCM.

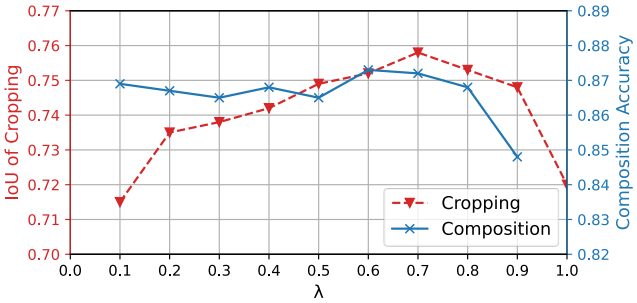

Figure 2. Sensitivity experiments of λ. The cropping performance reaches a peak when λ is set to 0.7. CACNet is generally robust to λ.

3. Sensitivity Experiments of λ

The hyperparameter λ is used to balance the cropping loss $\mathcal{L}_{crop}$ and the composition loss $\mathcal{L}_{com}$. We conduct sensitivity experiments of the balancing factor λ on the held-out validation set of FCDB. The results are shown in Fig. 2. One can observe that the best cropping performance can be achieved when λ is set to 0.7. Notably, the cropping performance degrades when the classification accuracy drops ($\lambda = 0.8$, $\lambda = 0.9$) even the cropping branch is dominant. Further, the cropping suffers substantial performance drop when we remove the composition branch ($\lambda = 1.0$). The results can verify the effectiveness of the composition branch to image cropping. Overall our method is robust to λ in cropping when it is set in the range of $[0.2, 0.9]$.

4. Further Qualitative Results

We present further qualitative results on two benchmarks FCDB [1] and FLMS [3], including:

- *Comparisons with the advanced methods,* i.e.*, VFN [2], VEN [8], VPN [8], A2RL [5], GAIC [9]*: In Fig. 3, our method generally generates more appealing cropping results close to the ground truth.
- *The interpretable image cropping*: Here we present the interpretable cropping of images that predicted obeying 2 or 3 rules. An image is regarded as obeying also a composition rule that yields a higher confidence score than a preset threshold 0.1. Fig. 4 and 5 shows the CAMs of the predicted top-2 or top-3 rules, the KCM, the weighted anchor points, the cropping result, and the composition distributions. With CAMs accurately localizing the discriminative regions of the corresponding rules, KCM further encodes the global composition evidences.
- *The KCM and cropping results of different composition rules*: Results are shown in Fig. 6. We do not provide results of the `pattern` rule because few images in natural scenes follow this rule. The dominant subjects, *e.g.*, the leading lines, the curves, and the geometric shapes, can be localized by KCM.

5. User Study

We design an online scoring website for the user study. The interface of the website is shown in Fig. 7, where 4 groups of images are illustrated. For the cropping results from different methods, we mix and pool them in a random order. Both the crop-out region and the cropped region are displayed simultaneously, and the crop-out region is covered with a darker mask. 15 visitors with photographing experience are invited to choose an option from 'Good', 'Normal', and 'Bad' for each cropping result. We count the number of the three options for each method and make comparisons.

References

[1] Yi-Ling Chen, Tzu-Wei Huang, Kai-Han Chang, Yu-Chen Tsai, Hwann-Tzong Chen, and Bing-Yu Chen. Quantitative analysis of automatic image cropping algorithms: A dataset and comparative study. In *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, pages 226–234, 2017. 2

[2] Yi-Ling Chen, Jan Klopp, Min Sun, Shao-Yi Chien, and Kwan-Liu Ma. Learning to compose with professional photographs on the web. In *Proc. ACM Int. Conf. Multimedia*, pages 37–45, 2017. 2

[3] Chen Fang, Zhe Lin, Radomir Mech, and Xiaohui Shen. Automatic image cropping using visual composition, boundary simplicity and content preservation models. In *Proc. ACM Int. Conf. Multimedia*, pages 1105–1108, 2014. 2

[4] Jun-Tae Lee, Han-Ul Kim, Chul Lee, and Chang-Su Kim. Photographic composition classification and dominant geometric element detection for outdoor scenes. *J. Vis. Commun. Image Represent.*, 55:91–105, 2018. 1

[5] Debang Li, Huikai Wu, Junge Zhang, and Kaiqi Huang. A2-rl: Aesthetics aware reinforcement learning for image cropping. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8193–8201, 2018. 2

[6] Debang Li, Junge Zhang, Kaiqi Huang, and Ming-Hsuan Yang. Composing good shots by exploiting mutual relations. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4213–4222, 2020. 1

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proc. Int. Conf. Learn. Represent.*, 2015. 1

[8] Zijun Wei, Jianming Zhang, Xiaohui Shen, Zhe Lin, Radomir Mech, Minh Hoai, and Dimitris Samaras. Good view hunting: Learning photo composition from dense view pairs. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5437–5446, 2018. 1, 2

[9] Hui Zeng, Lida Li, Zisheng Cao, and Lei Zhang. Reliable and efficient image cropping: A grid anchor based approach. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5949–5957, 2019. 1, 2

Figure 3. Further qualitative comparison of different methods.

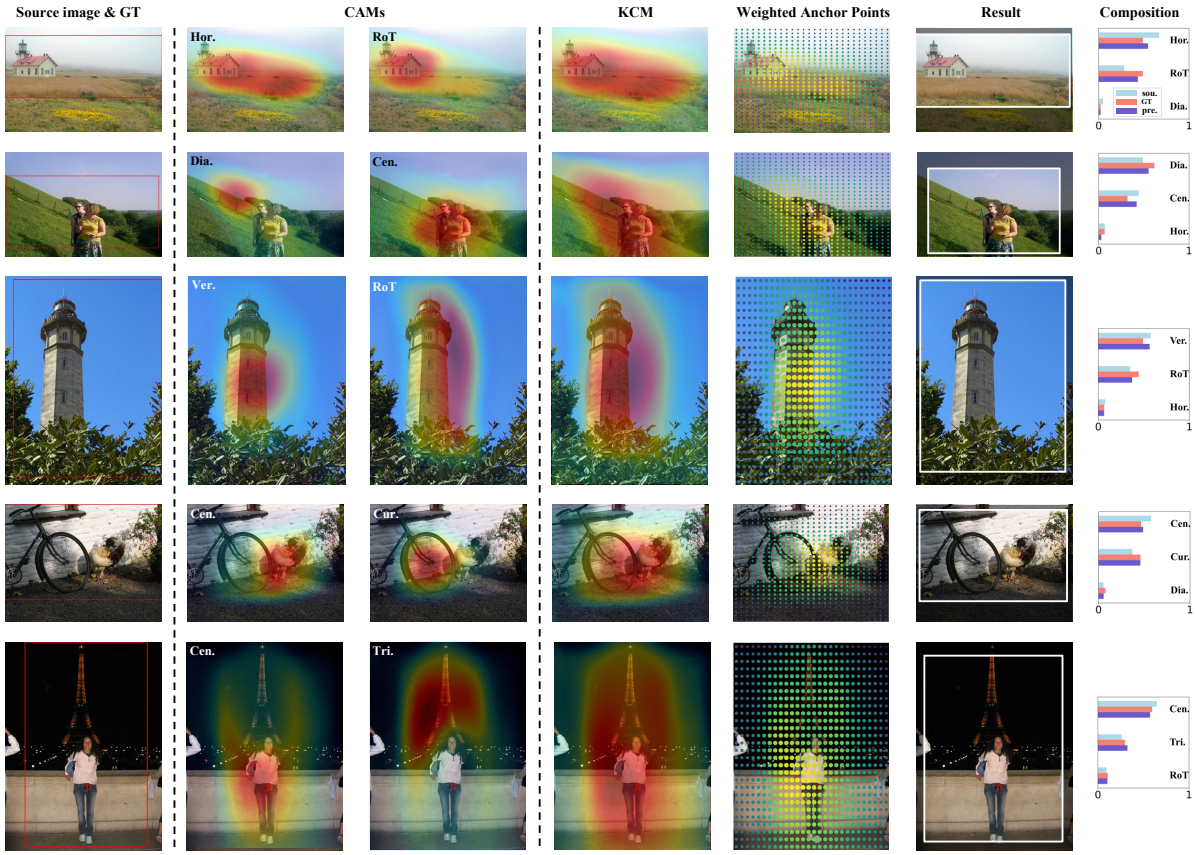

Figure 4. Interpretable image cropping of images in 2 composition rules.

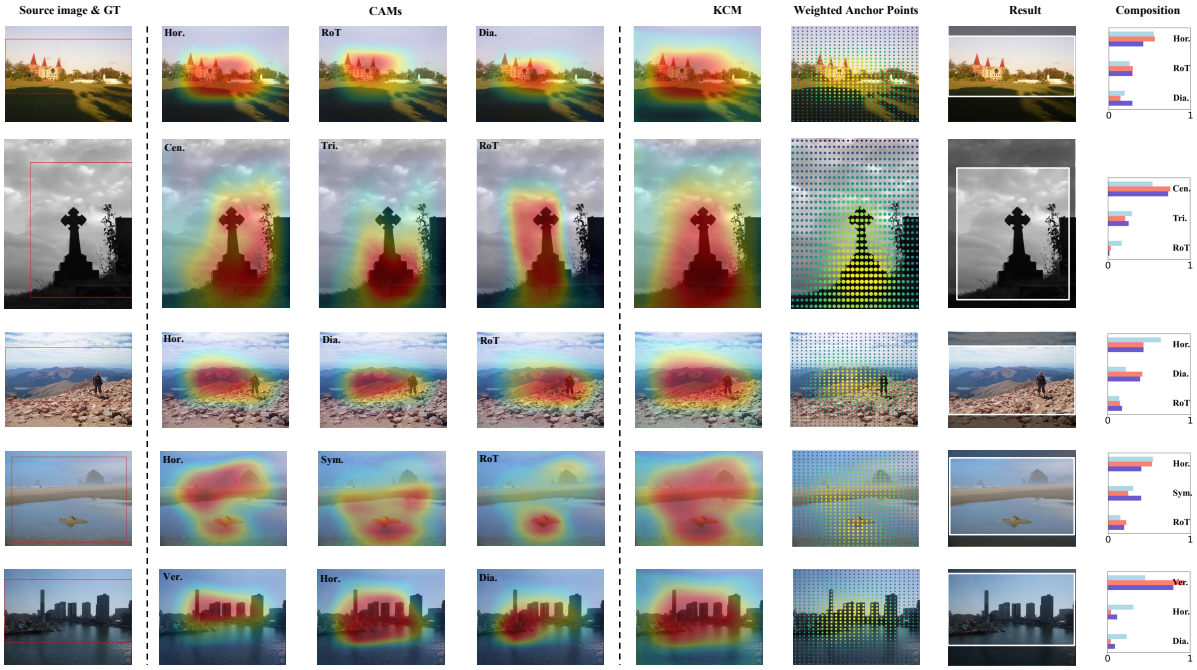

Figure 5. Interpretable image cropping of images in 3 composition rules.

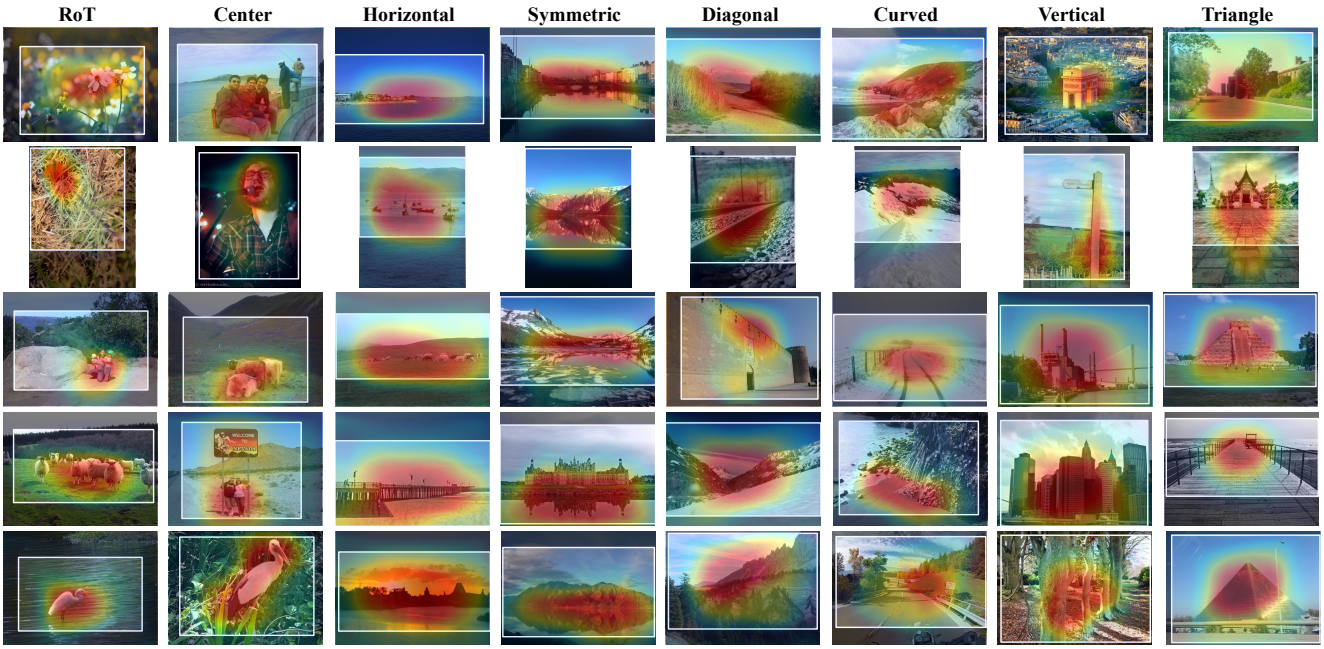

Figure 6. Further visual results of the interpretable cropping of different composition rules. With the KCM encoding the global composition, our method produces well-composed results of each rule.

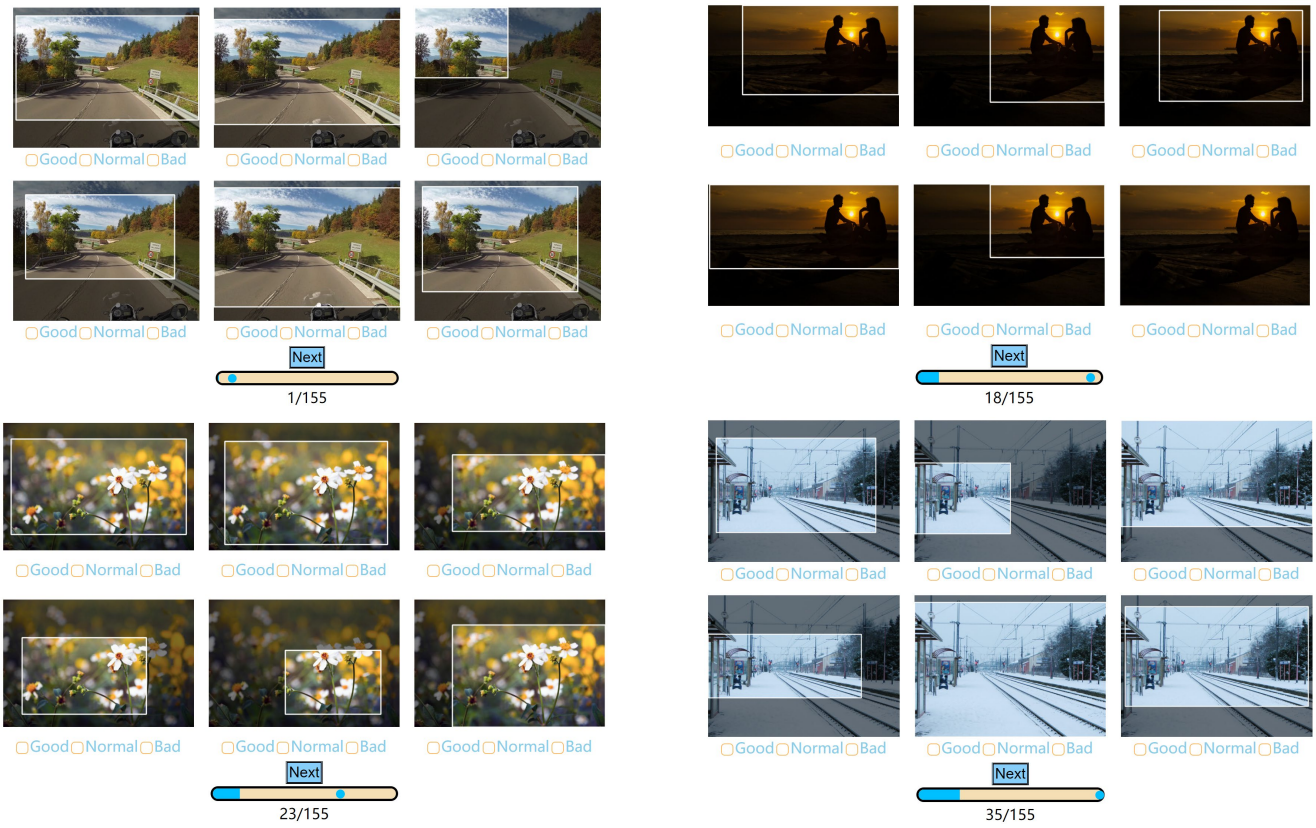

Figure 7. Design of the user study interface.