

Content-Based Visual Summarization for Image Collections

Xingjia Pan, Fan Tang, Weiming Dong, *Member, IEEE*, Chongyang Ma, Yiping Meng, Feiyue Huang, Tong-Yee Lee, *Senior Member, IEEE*, Changsheng Xu, *Fellow, IEEE*

1 TECHNICAL DETAILS ON TREE-BASED LAYOUT

1.1 Layout Optimization

Coarse Adjustment Intuitively, the way of horizontal division makes the aspect ratios of subregions smaller as the width of subregions remain the same while the height become smaller. Similarly, the column division makes subregions more narrow. Based on the splitting way, we adjust the division ratio α next. If adjusted splitting way is H ,

$$\alpha = lc_ar / (lc_ar + rc_ar) \quad (1)$$

while if splitting is V ,

$$\alpha = rc_ar / (lc_ar + rc_ar) \quad (2)$$

where lc_ar and rc_ar are the real aspect ratio of two images associating with two child nodes respectively. Considering the ratio is based on the real child nodes size, we get

$$\alpha_{sz} = lc_sz / (lc_sz + rc_sz) \quad (3)$$

where lc_sz and rc_sz are the real size of two images corresponding to two child nodes respectively. α and α_{sz} are different usually. To bridge the gap of these two parameters, we reassign α

$$\alpha = \lambda_a * \alpha + \lambda_s * \alpha_{sz} \quad (4)$$

where λ_a, λ_s are the weight parameters. Then the sz_{exp} is adjusted correspondingly,

$$\begin{cases} lc_sz_{exp} = sz_{exp} * \alpha \\ rc_sz_{exp} = sz_{exp} * (1.0 - \alpha) \end{cases} \quad (5)$$

- X. Pan, W. Dong and C. Xu are with NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China and University of Chinese Academy of Sciences, Beijing, China. E-mail: {panxingjia2015, weiming.dong, changsheng.xu}@ia.ac.cn.
- F. Tang is with Fosafer, Beijing, China. E-mail: tangfan@fosafer.com.
- C. Ma is with Kuaishou Technology, Beijing, China. E-mail: chongyangm@gmail.com.
- Y. Meng is with Didi Chuxing, Beijing, China. E-mail: mengyipingkitty@didichuxing.com.
- F. Huang is with YouTu Lab, Tencent, Shanghai, China. E-mail: gary.huang@tencent.com.
- T.-Y. Lee is with National Cheng Kung University, Tainan, Taiwan. E-mail: tonylee@mail.ncku.edu.tw.

where lc_sz_{exp} and rc_sz_{exp} are the expected size of child nodes respectively. Similarly, the ar_{exp} is assigned,

$$lc_ar_{exp} = \begin{cases} ar_{exp} * \alpha, & S = ' H' \\ ar_{exp}/\alpha, & S = ' V' \end{cases} \quad (6)$$

$$rc_ar_{exp} = \begin{cases} ar_{exp} * (1.0 - \alpha), & S = ' H' \\ ar_{exp}/(1.0 - \alpha), & S = ' V' \end{cases} \quad (7)$$

where lc_ar_{exp} and rc_ar_{exp} are the expected aspect ratio of child nodes respectively.

Considering the number of images associated to a node make great influence to node size, we adjust the N_{exp} with different strategy from ones of sz_{exp} and ar_{exp} , that is, assigning N_{exp} according the α_{sz} mentioned above,

$$\begin{cases} lc_N_{exp} = N_{exp} * \alpha_{sz} \\ rc_N_{exp} = N_{exp} * (1.0 - \alpha_{sz}) \end{cases} \quad (8)$$

here lc_N_{exp} and rc_N_{exp} are the expected number of images associated to child nodes respectively.

In the process of adjustment, we truncate the subtree where the N_{exp} and N of node are different and then regenerate the subtree as expected according to $N_{exp}, ar_{exp}, sz_{exp}$. After one adjustment through a tree, we reassigned units to each anchor node, images to each leaf node by Hungarian Algorithm and then to another iteration until meet the termination criteria. Corresponding to the two-stage constructed tree, we alternately adjust the anchor node and leaf node, namely, adjusting one with fixing another one.

Refined Optimization Given the aspect ratio and width of collage, we can calculate the width and aspect ratio of each cell associated to one image according to the splitting way, S , and splitting ratio, α , of each node. Thus, to optimize the collage further, we adjust α of each node delicately. Meanwhile, with cell width, w , and α , the cell height is determined accordingly, that is, the size of cell depend on α and w . The issue of placing image content-aware and aesthetically is achieved through Hungarian algorithm, so our refined optimization mainly focuses on the other two issues. To get the ar_{exp} of left child nodes, we according to

$$f_l(\alpha, S | ar_{exp_i}) = \begin{cases} ar_{exp_i} * \alpha, & S = ' V' \\ ar_{exp_i}/\alpha, & S = ' H' \end{cases} \quad (9)$$

TABLE 1
Participant statistic.

Evaluation	male/female	age range
P1-P15	7/8	18-30
P16-P30	8/7	17-28
P31-P50	15/15	17-35

similarly, to get right child nodes' ar_{exp} ,

$$f_r(\alpha, S | ar_{exp_i}) = \begin{cases} ar_{exp_i} * (1 - \alpha), & S = V' \\ ar_{exp_i} / (1 - \alpha), & S = H' \end{cases} \quad (10)$$

For convenience, we define F_i ,

$$F_i(\alpha, S | ar_{exp_i}) = \begin{cases} f_l(\alpha, S | ar_{exp_i}), & lc \\ f_r(\alpha, S | ar_{exp_i}), & rc \end{cases} \quad (11)$$

where i indicates the depth of the node. Similarly, to get child node width, w_{exp} , we define H_i ,

$$H_i(\alpha, S | w_{exp_i}) = \begin{cases} ph_l(\alpha, S | w_{exp_i}), & lc \\ hr_r(\alpha, S | w_{exp_i}), & rc \end{cases} \quad (12)$$

where w_{exp_i} is the expected width of node in i^{th} layer. To utilize back propagation algorithm, we need to calculate the partial derivative of each node by using chain rule. To calculate the partial derivative of C_{sh} and C_{sz} with respect to ar_{exp} of node in depth i , respectively,

$$\begin{aligned} \frac{\partial C_{sh}}{\partial ar_{exp}^i} &= \frac{\partial C_{sh}}{\partial ar_{exp}} * \frac{\partial F_n}{\partial F_{n-1}} * \frac{\partial F_{n-1}}{\partial F_{n-2}} * \dots * \frac{\partial F_{i+1}}{\partial ar_{exp}^i} \\ \frac{\partial C_{sz}}{\partial ar_{exp}^i} &= \frac{\partial C_{sz}}{\partial ar_{exp}} * \frac{\partial H_n}{\partial H_{n-1}} * \frac{\partial H_{n-1}}{\partial H_{n-2}} * \dots * \frac{\partial H_{i+1}}{\partial ar_{exp}^i} \end{aligned} \quad (13)$$

where ar_{exp}^i is the ar_{exp} of the node in depth i . Thus, considering both the issue 1) and 2) synthetically, to get the arbitrary partial derivative of cost function, C_{sh} and C_{sz} , with respect to α in depth i ,

$$\begin{aligned} \frac{\partial(C_{sh} + C_{sz})}{\partial \alpha^i} &= \frac{\partial C_{sh}}{\partial ar_{exp}} * \frac{\partial F_n}{\partial F_{n-1}} * \frac{\partial F_{n-1}}{\partial F_{n-2}} * \dots * \frac{\partial F_{i+1}}{\partial \alpha^i} \\ &\quad + \frac{\partial C_{sz}}{\partial ar_{exp}} * \frac{\partial H_n}{\partial H_{n-1}} * \frac{\partial H_{n-1}}{\partial H_{n-2}} * \dots * \frac{\partial H_{i+1}}{\partial \alpha^i} \\ &= \frac{\partial C_{sh}}{\partial ar_{exp}^{i+1}} * \frac{\partial F_{i+1}}{\partial \alpha^i} + \frac{\partial C_{sz}}{\partial ar_{exp}^{i+1}} * \frac{\partial H_{i+1}}{\partial \alpha^i} \end{aligned} \quad (14)$$

Based on the gradients of α of each node, we use the 'momentum update' referring to deep learning as well to update the α .

2 EXPERIMENTAL DETAILS AND MORE RESULTS

2.1 Photo summarization

In our experiments, to overcome the bias of different software preferences, we have some settings as followed: the cluster number used in K-means clustering is the same as that of the representative images selected by the diversity gain constraint; the image overlapping rate is set to the minimum value and the number of images selected automatically is set to the summarization size in AutoCollage; the parameters of both tools are set as default unless otherwise specified. We here show some more results of photo summarization by random selection

(without replacement), k-means clustering, k-means-D and our HTS method respectively in Figs. 2,3 and 4. K-means-D means that we select the most diverse image as the represent one from each cluster based on our Div while K-means select the closest K images to the cluster centers. Tab. 1 shows the statistics of participants involved in our userstudy.

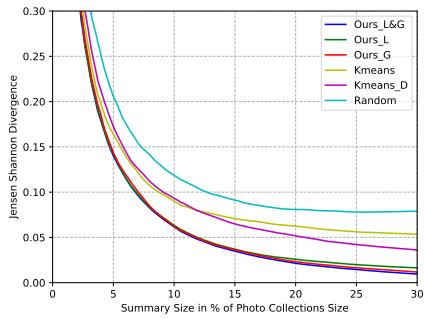
In addition, we conduct more experiments than the main paper. We calculate $\delta(\cdot)$ based on the features of ImageNet and Places365 validation dataset as the global item, called $\delta(\cdot)_G$, as well as the original gallery, called $\delta(\cdot)_L$, as the local item, respectively. Thus the summarization results by our method are called $Ours_L$, $Ours_G$ and $Ours_{L\&G}$ with the disparity item $\delta(\cdot)_G$, $\delta(\cdot)_L$ and both, respectively. Note that all the results are averaged cross all galleries. The results are presented in Fig. 1. The x-axis represents summarization size and the y-axis present the JS divergency, reconstruction error, diversity ratio (as fraction of original dataset) and Rep defined in Sec. 4 of main paper, respectively. In all the cases, our three results outperforms the baselines. The summarization using $\delta(\cdot)_G$ performs best for it excludes the local noise of the original gallery. In addition, K-means-D, selecting each image from every cluster based on our Div , outperforms the conventional k -means clustering, which indicates our Div improve the informativeness of the summarization. Furthermore, although the K-means clustering implicitly minimize the reconstruct error in mechanism, our method also performs better.

2.2 Photo Collage

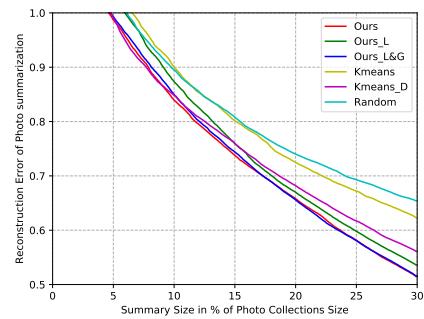
We here show the collages in main paper again and some more results in Figs. 5-12.

REFERENCES

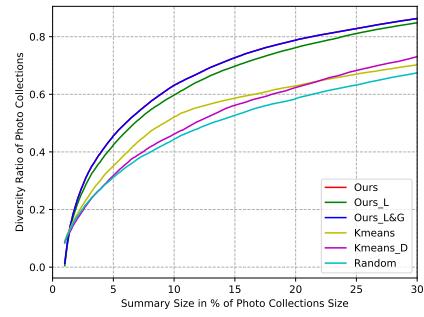
- [1] V. Cheung, "Shape collage," Available: <https://shapecollage.com/>, 2013, online.
- [2] T. Liu, J. Wang, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum, "Picture collage," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1225–1239, 2009.
- [3] Y. Liang, X. Wang, S. H. Zhang, S. M. Hu, and S. Liu, "Photorecomposer: Interactive photo recomposition by cropping," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 10, pp. 2728–2742, Oct 2018.



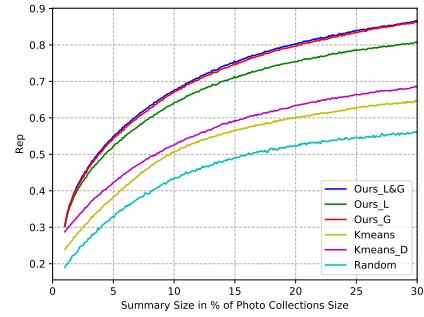
(a) JS Divergence



(b) Reconstruction Error



(c) Diversity Ratio



(d) Rep

Fig. 1. Quantitative results of photo summarization.



Fig. 2. Summarization of Artem Beliaikin's gallery. From top to down are summarizations by *random selection*, *k-means*, *k-means-D* and ours respectively.

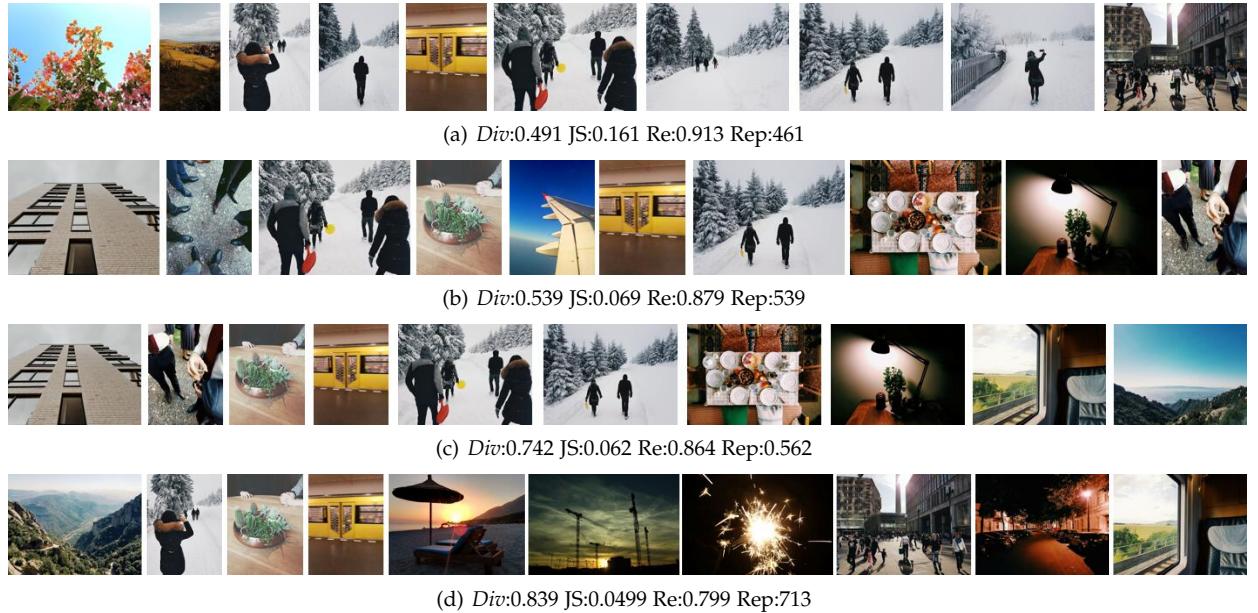


Fig. 3. Summarization of Daniel Frese's gallery. From top to down are summarizations by *random selection*, *k-means*, *k-means-D* and ours respectively.

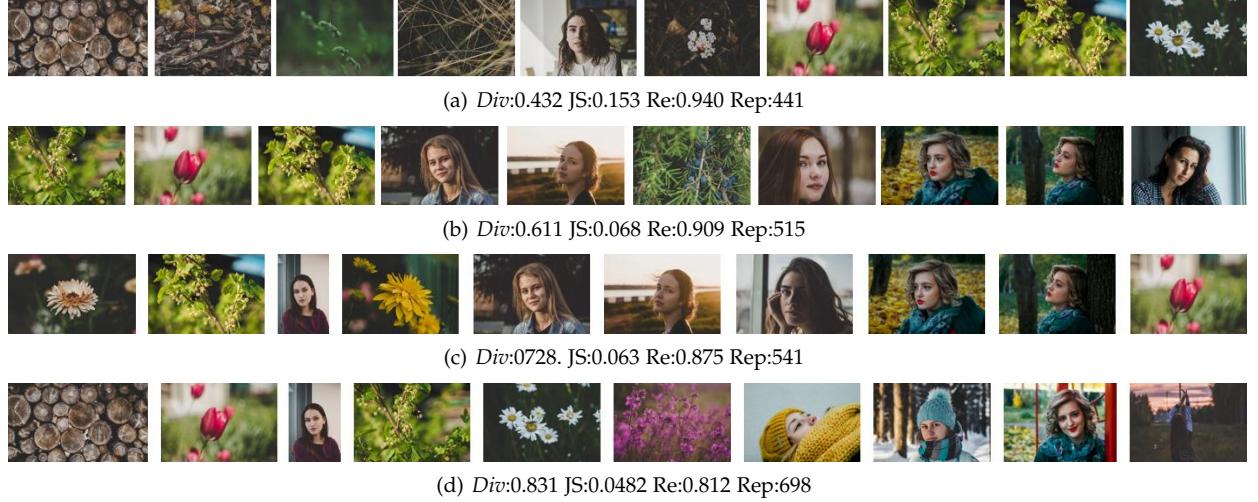


Fig. 4. Summarization of Daniel Spase's gallery. From top to down are summarizations by *random selection*, *k-means*, *k-means-D* and ours respectively.

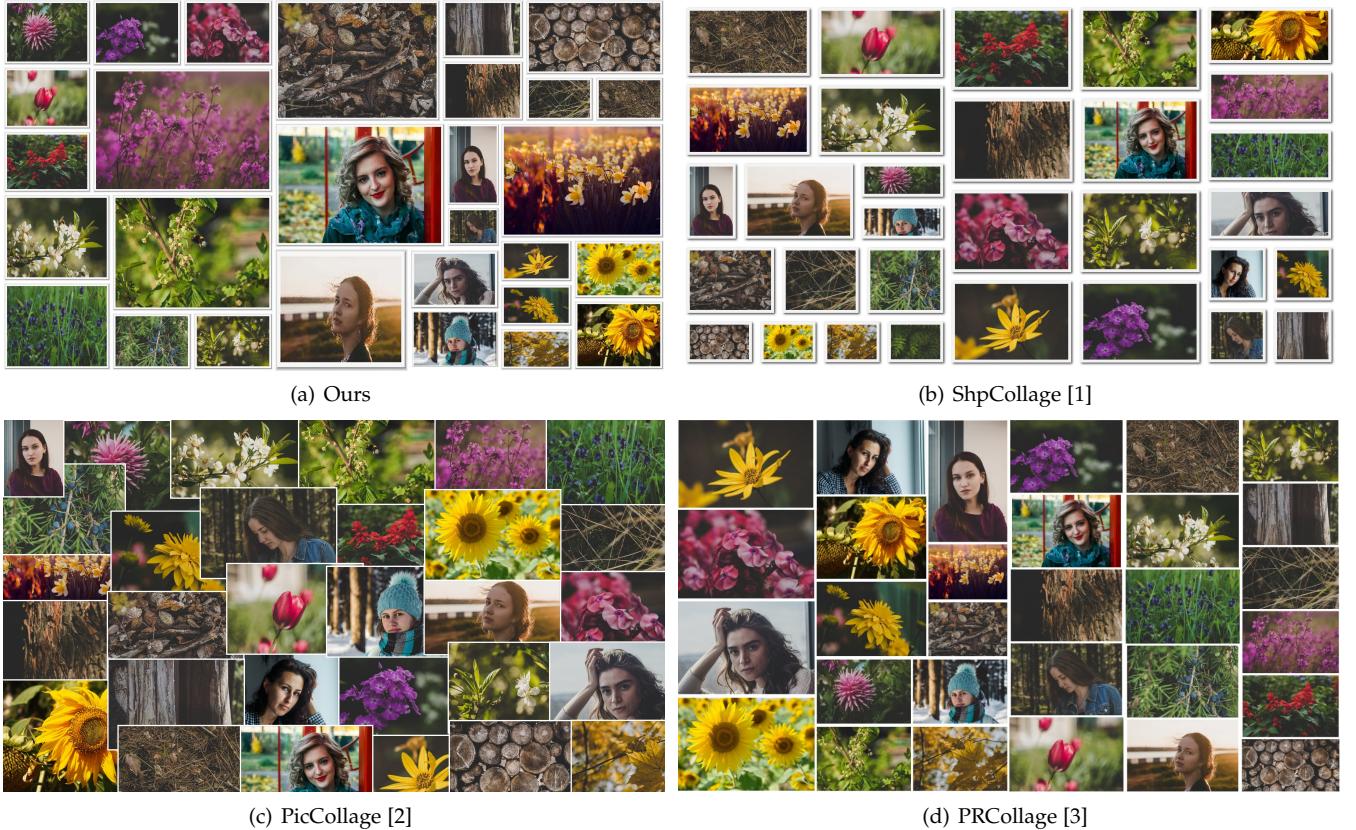
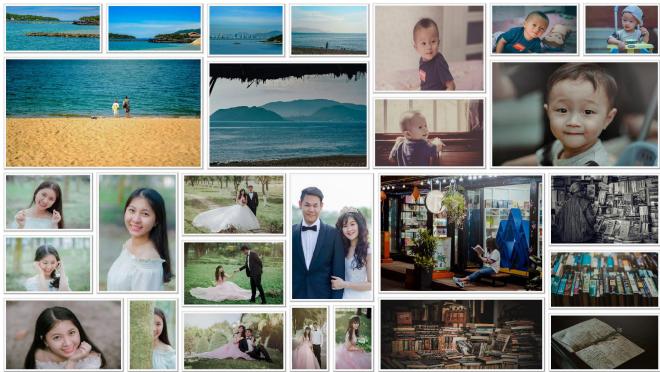


Fig. 5. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.



(a) Ours



(b) ShpCollage [1]



(c) PicCollage [2]



(d) PRCollage [3]

Fig. 6. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.



(a) Ours



(b) ShpCollage [1]



(c) PicCollage [2]



(d) PRCollage [3]

Fig. 7. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.

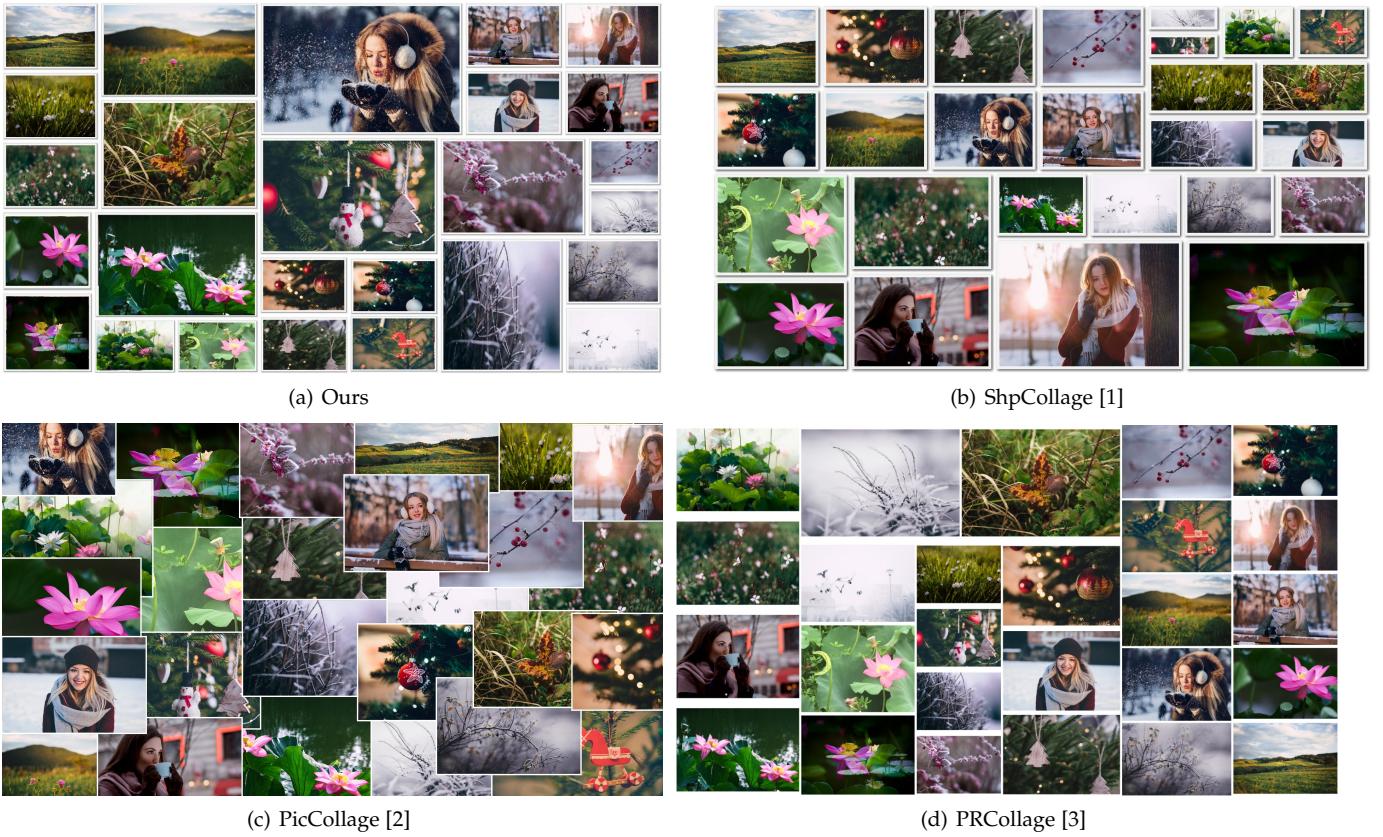


Fig. 8. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.

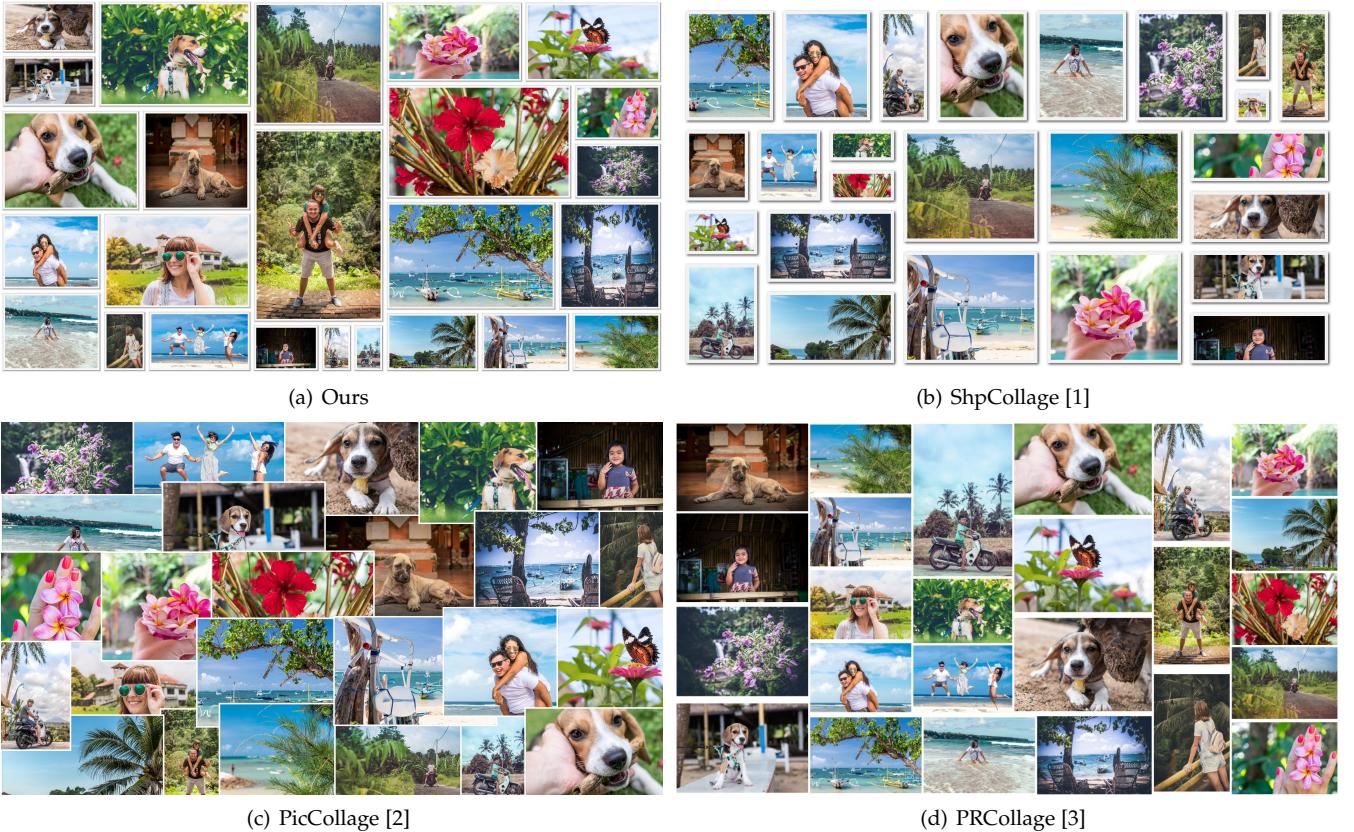


Fig. 9. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.



Fig. 10. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.

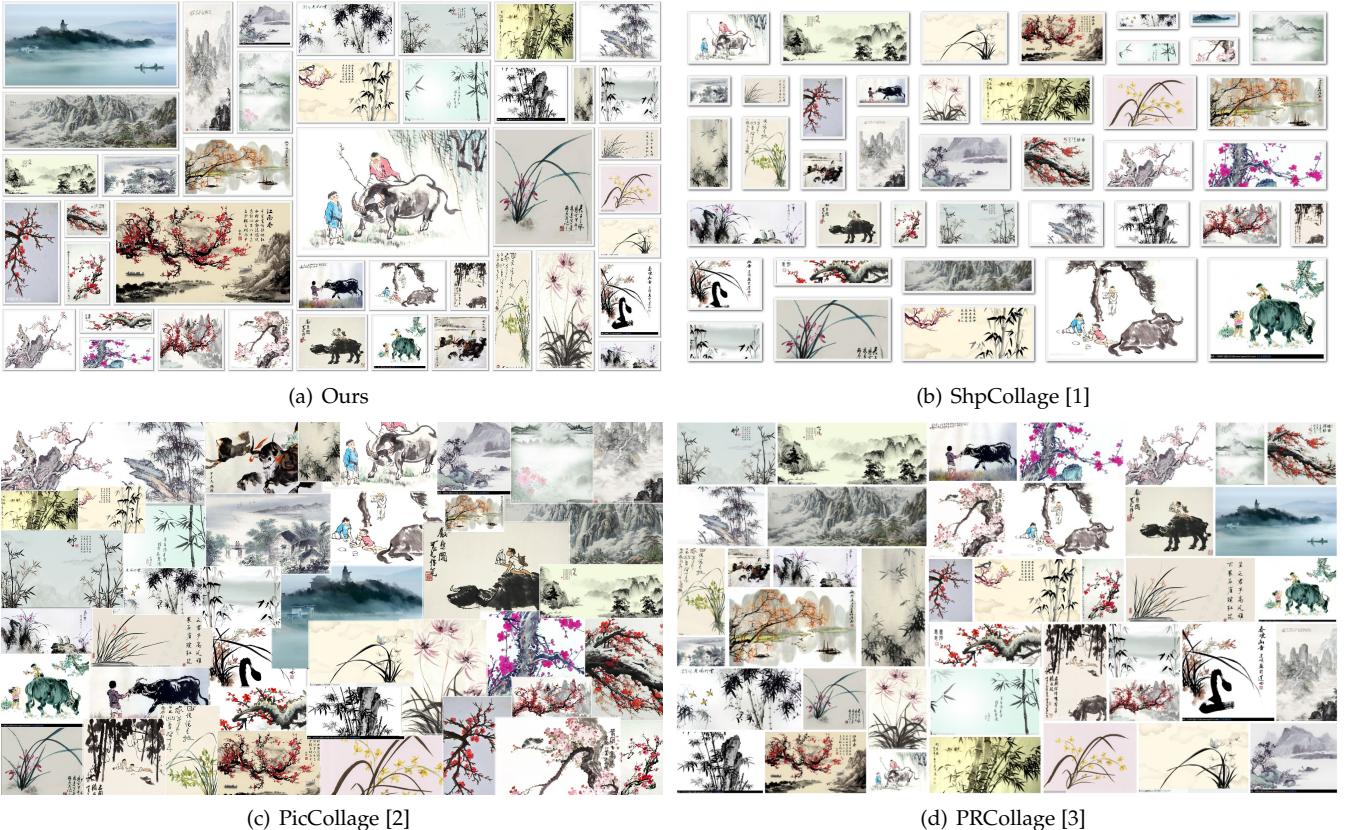


Fig. 11. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.



Fig. 12. Comparison with other methods. From left to right, top to down are Ours, ShpCollage [1], Picture Collage [2], PRCollage [3] respectively.

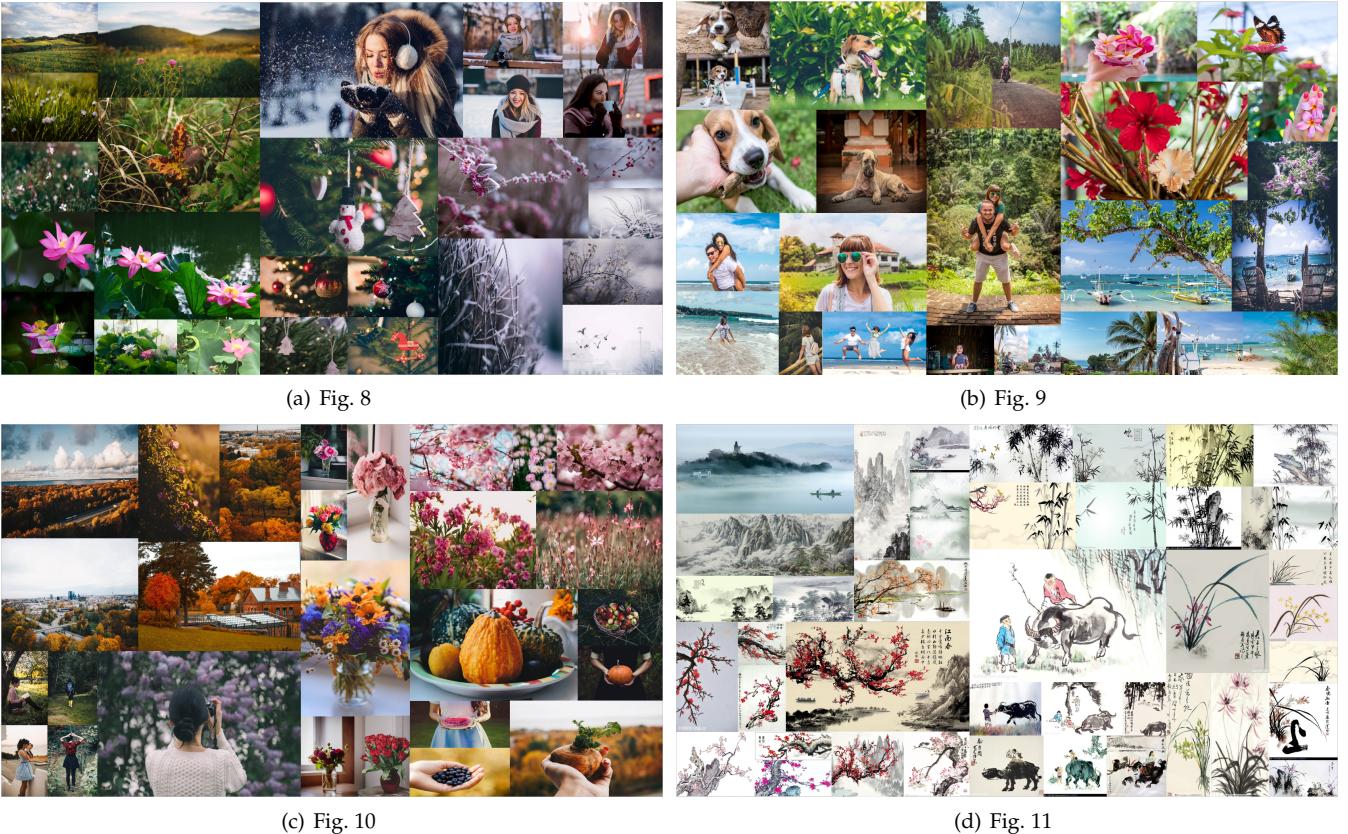


Fig. 13. Our collages without white space for Figs. 8 and 11, respectively.



Fig. 14. Our collages without white space for Fig. 12.