Deep Feature Guided Image Retargeting

Jinan Wu[1], Rong Xie[1], Li Song[1,2], Bo Liu[3]
[1]Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
[2]MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai 200240, China
[3]School of computer science, University of Technology Sydney, Australia
{13414207, xierong, song_li}@sjtu.edu.cn, bo.liu@uts.edu.au

***Abstract*—Image retargeting is the technique to display images via devices with various aspect ratios and sizes. Traditional content-aware retargeting methods rely on low-level features to predict pixel-wise importance and can hardly preserve both the structure lines and salient regions of the source image. To address this problem, we propose a novel adaptive image warping approach which integrates with deep convolutional neural network. In the proposed method, a visual importance map and a foreground mask map are generated by a pre-trained network. The two maps and other constraints guide the warping process to yield retargeted results with less distortions. Extensive experiments in terms of visual quality and a user study are carried out on the widely used RetargetMe dataset. Experimental results show that our method outperforms current state-of-art image retargeting methods.**

***Index Terms*—Image retargeting, CNN, image warping.**

I. Introduction

With the increasing diversity and versatility of display devices, image retargeting techniques which adapt the image to different resolutions and aspect ratios are becoming an active research topic. The naive methods like linear scaling or manual cropping may suffer from severe information loss and distortions. Thus, many content-aware image retargeting methods have been proposed in the past decades. The main objective of content-aware image retargeting is to change the aspect ratio of the original images arbitrarily while maintaining the appearance of salient objects, as well as reducing artifacts.

Content-aware image retargeting uses a visual importance map to identify the important regions. Most existing retargeting methods such as seam carving [1] and scale-and-stretch [2], use either a gradient map or saliency detection to generate an importance map which indicates pixel-wise importance. However, such importance map mainly measures account for low-level features, such as edges, color contrast, or handcrafted features. Its generality is limited due to its lack of high-level semantic features. In recent years, convolutional neural network (CNN) has demonstrated outstanding performance in multi-level computer vision fields. T herefore, h ow t o apply deep learning technique in image retargeting needs to be further explored.

In this paper, we combine CNN with an image warping method to take semantic features into account. There are two major contributions in this paper: First, we propose a novel importance map that is generated by fusing the intermediate output of a foreground object segmentation network. Second, we improve the image warping algorithms by applying

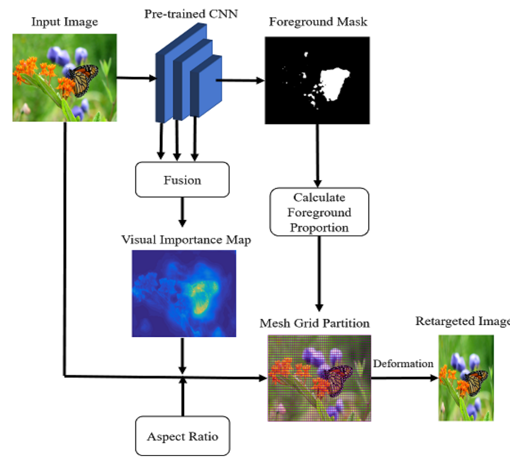

Fig. 1. Overview of the deep feature guided image retargeting scheme.

adaptive grids generation. Experimental results show that our proposed retargeting approach yield more visually pleasing results than other conventional approaches.

The rest of the paper is organized as follows. We introduce related work in Section II and present the details of our method in Section III. The experimental results and analysis are shown in section IV. Our conclusion and future work are presented in Section V.

II. Related Work

Numerous content-aware image retargeting techniques have been carried out in the field of image processing. We can roughly classify them into two main categories: discrete and continuous. The discrete approaches manipulate pixels at unimportant and homogenous areas to change the aspect ratio of an image. Seam carving [1] performs retargeting by iteratively inserting or removing an optimal seam via dynamic programing. However, it performs poorly when salient content is non-textured. Pritch *et al.* [3] proposed shift-map for pixel re-arrangement, which can be applied in various image editing applications, including retargeting. The continuous approaches seek an optimal mapping or warping through constraining deformation and smoothness. Wolf *et al.* [4] proposed a warping function by non-homogeneously squeezing the image in one direction. Wang *et al.* [2] presented an optimized scale-and-stretch warping method, which iteratively warps local regions to match the optimal scaling factors as close as possible.

To achieve better performance in both texture and salient object preservation, the importance map is the multiplication of the edge map and saliency map. However, inconsistent deformations of prominent objects occupying several quads may still occur.

Visual significance map is crucial for most content-aware image retargeting operators, and different maps may lead to totally different retargeted results. To address this problem, Cho $et\ al.$ [5] first applied a weakly- and self-supervised deep CNN model in image retargeting. The authors design an encoder-decoder network which produces a semantic-aware shift map in order to obtain better target image. However, the size of the input image of this network is strictly limited, otherwise it will lead to serious artifacts. Lin $et\ al.$ [6] proposed a method named DeepIR which utilizes standard classification network and performs retargeting in feature space. However, its time complexity is high due to its image reconstruction procedure. In this work, we utilize a segmentation-oriented CNN to extract semantic information from the input image, and conduct warping process under its guide. Our method shows its superiority in some challenging scenes and achieves comparable results in a limited time.

III. Proposed Method

The overall scheme of our deep feature guided image retargeting method is illustrated in Fig. 1.

A. Semantic-aware Importance Map Generation

In order to leverage the high-level features of the input image, unlike previous works [6] which use networks trained for classification (i.e.,VGG19 [7]), we utilize a network trained for foreground object segmentation proposed in [8] which architecture largely follows the standard VGG16 architecture. Assume we use the L layers of the pre-trained CNN, when we feed an input image I of size $(h \times w)$ to it and perform a forward propagation, we could get a set of feature maps of the input image:

$$\Phi(I) = \{\Phi(I)_1, \Phi(I)_2, ..., \Phi(I)_L\}. \quad (1)$$

The feature map denoted as $\Phi(I)_i$ is 3-D tensors with size $(h_i \times w_i \times c_i)$, where h is height, w is width and c is channel number.

It has been proven in [9] that region of higher semantic significance will result in stronger activation in feature maps. The importance map of layer i is defined as:

$$S_i(I)(i,j) = ||\sum_{c=1}^{c_i} \Phi(I)(i,j,c)||_2. \quad (2)$$

In order to eliminate layer-dependent amplitude differences, we normalize S_i into $\bar{S}_i$ by min-max normalization. Before the fusing step, we resize $\bar{S}_i$ into the same size as the input image.

Finally, taking both the low-level and high-level features into account, we fuse the importance map of each layer as follows:

$$S(I)(i,j) = \sum_{i=1}^{L} \alpha_i \times \bar{S}_i(I)(i,j). \quad (3)$$

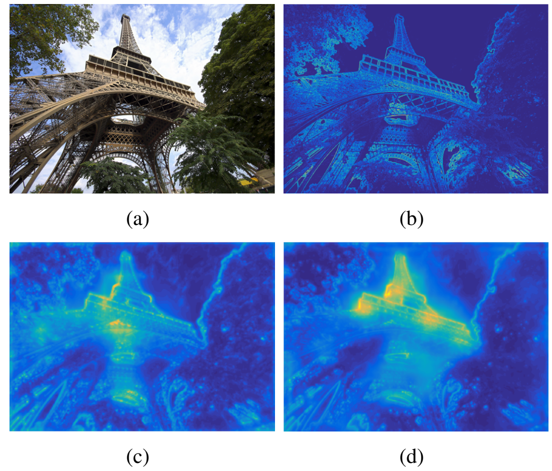

Fig. 2. Comparision of different importance maps. (a)original image; (b)gradient map; (c)importance map by fusing feature maps of standard VGG19; (d)ours.

As is shown in Fig. 2., the proposed importance map performs better to concentrate on salient regions than gradient map and importance map by fusing feature maps of VGG19 trained for classification. Besides, through this progressive accumulation, our importance map is smooth enough to propagate the distortions and avoid wrong segmention result, which will be discussed in section IV.

B. Adaptive Grid Partition

For most continuous image warping, a quad mesh **M=(V,E,Q)** containing a vertex set **V**, an edge set **E**, and a quad face set **Q** are created for image warping at first. The significance w_f of quad f is defined as the average pixel significance within the quad. Obviously, the resolution of the quads is a crucial parameter. The higher the quad resolution we set, the higher the line preserving quality but the lower the computation efficiency and deformation ability we obtain. The quad resolution of previous work [2], [10] is usually set to 20×20 pixels to make a trade-off.

We find that retargeted image is not very sensitive to the quad resolution unless the input image is a scene image without any salient object. The pre-trained CNN we use is able to produce a high-quality foreground object segmentation map which helps us to classify the input image roughly. We calculate proportion of the foreground area to the original picture and denote it as F_p. The grid size is decided as:

$$grid\ size = \begin{cases} 20, F_p \leq 10\%; \\ 30, 10\% < F_p \leq 50\%; \\ 40, F_p > 50\%. \end{cases} \quad (4)$$

For our method, 40×40 pixels is sufficient to retarget images efficiently. When the salient regions occupy almost the whole input image, lower quad resolution enables warping process to squeeze unimportant background better.

C. Image Warping

This section is regarded as the grid deformation procedure. Assume that the original image of $m \times n$ pixels is resized to the size of $m' \times n'$. The proposed warping method spread the distortion according to the significance of each quad and obtain the new mesh geometry **V'** by a global optimization to change the positions of the vertices in **V**.

We adapt the energy function which is proposed in [2]. To conduct similar transform of prominent regions and alleviate structure distortions, quad deformation energy and grid line bending energy, are introduced to the during the optimization. Ideally, for each $f \in \boldsymbol{F}$ there would be a scale factor s_f such that for each vertex *v* of the quad, **v**' = s_f**v** + **t** (where **t** is a constant translation vector). The set of (directed) edges of *f* is denoted as **E**(f); the total quad deformatiom energy is defined as:

$$D_u = \sum_{f \in \mathbf{F}} w_f \times \sum_{\{i,j\} \in \mathbf{E}(f)} ||(v_i' - v_j') - s_f(v_i - v_j)||^2 \quad (5)$$

To prevent distortion of the relatively large objects, bending of grid lines needs to be punished. The length ratio of the edges before and after deformation as $l_{ij} = ||(v_i' - v_j')||/(v_i - v_j)||$. The grid line bending energy is introduced as:

$$D_l = \sum_{\{i,j\} \in \mathbf{E}(f)} ||(v_i' - v_j') - l_{ij}(v_i - v_j)||^2 \quad (6)$$

Total energy is the sum of these two energy terms $D = 0.9D_u + 1.1D_l$. In this work, the solver starts with a uniform scaled quad mesh of deformed vertices **V**. Then the vertex set **V** is updated by solving a constrained least-squares system. The optimization procedure repeats until all the vertex movements are smaller than 0.5. Finally we warp the original image to desired size according to the obtained mesh geometry.

IV. Experiments And Results

A. Experimental Settings

We have implemented the proposed method with MATLAB and Pytorch on a PC equipped with Intel Core i7-7700 CPU 3.60GHz and NVIDIA TitanXP GPU. For a 1024×768image, the average computation time for importance map generation is 0.083s, and that for image warping is 0.052s.

The pre-trained CNN of our approach is trained on PASCAL 2012 [11]. To achieve a trade-off between high-level semantic content and low-level details, the value of parameters $\{\alpha_i\}$ used in Eq. (3) is set to {0.2,0.1,0.1,0.3,0.3}.We apply the proposed approach to the RetargeMe dataset [12] for qualitative and quantitative evaluation.

B. Evaluation and Analysis

The pre-trained CNN we adapt is one of the state-of-art foreground object segmentation method. Howerver, structure distortions may occur when we feed the foreground segmentation map into warping operator directly, which is shown in Fig. 3. It is obvious that a binarized segmentation map is not suitable for retargeting tasks. The distortions can be eased by our smooth importance map.

(a) (b) (c)

Fig. 3. Comparision of retargeted results by two importance maps(foreground segmentation map and ours). (a)original image; (b)results with foreground segmentation map; (c)ours.

According to previous comparative study [12], seam carving (SC) [1] and streaming video (SV) [13] are considered as the best discrete and continous retargeting methods, respectively. So in this experiment, five methods have been compared: naive retargeting approaches including linear scaling (SCL) and cropping (CR), SC, SV, and our algorithm. Qualitative results are shown in Fig. 4. It can be observed that the cropping method usually causes severe information loss. SCL method results in over reducing salient objects. SC tends to deform important objects when the salient region is non-texured. SV which relies on both human labeled and automatic features may cause over-stretching of the background. In general, our method is able to produce high-quality retargeted images without information loss automatically in various cases.

C. User Study

There is currently no recognized objective evaluation criteria for image retargeting. To evaluate our method further, a user study with 47 participants (25 males, 22 females, age range 20-49) from different backgrounds is conducted. Each participant is offered 60 sets of images which contain the original image and the resized images form the five methods. The images are selected form the RetargetMe [12] database of 0.75 or 0.5 aspect ratio. Then parcitipants choose their favourite retargetd image within 30 minutes. Skipping image sets which are hard to decide is allowed. A total of 2573 votes are collected, and the statistic result is shown in Fig. 5. Based on the statistics, our approach can potentially generate better retargeting results compared with the related approaches.

V. Conclusion And future work

In this paper, we present a novel deep feature guided image retargeting method. Our improved importance map which contains both the gradient information and semantic content is effective and smooth in distortions propagation. Adaptive grid size is selected based on the proportion of the foreground to adjust the deformation ability. These procedures enable the final mesh grid warping to ease artifacts and yield visually

Fig. 4. Half width reducing comparisons with CR, SCL, SC, SV and our method on the RetargetMe dataset.

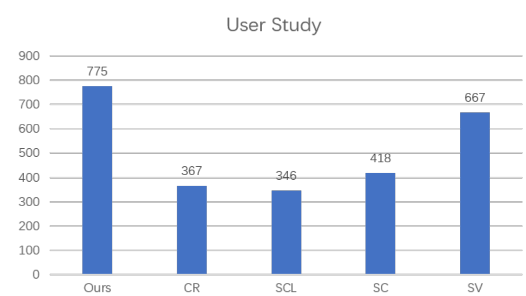

Fig. 5. The number of votes for the 5 methods considered in our user-study.

pleasing retargeted images. Our framework is flexible and can be easily combined with other retargeting methods for future works. The conducted comparisons and user study show that our proposed approach performs better than the current state-of-art methods. In the future, a fine-grained semantic segmention is considerd to be integrated with our method.

Acknowledgment

This work was supported by MoE-China Mobile Research Fund Project (MCM20180702), NSFC (U1611461, 61521062, 61671296), the 111 Project (B07022 and Sheitc No.150633) and the Shanghai Key Laboratory of Digital Media Processing and Transmissions.

References

[1] Shai Avidan and Ariel Shamir, "Seam carving for content-aware image resizing," in *ACM Transactions on graphics (TOG)*. ACM, 2007, vol. 26, p. 10.

[2] Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee, "Optimized scale-and-stretch for image resizing," in *ACM Transactions on Graphics (TOG)*. ACM, 2008, vol. 27, p. 118.

[3] Yael Pritch, Eitam Kav-Venaki, and Shmuel Peleg, "Shift-map image editing," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 151–158.

[4] Lior Wolf, Moshe Guttmann, and Daniel Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–6.

[5] Donghyeon Cho, Jinsun Park, Tae-Hyun Oh, Yu-Wing Tai, and In So Kweon, "Weakly-and self-supervised learning for content-aware deep image retargeting," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4558–4567.

[6] Jianxin Lin, Tiankuang Zhou, and Zhibo Chen, "Deepir: A deep semantics driven framework for image retargeting," *arXiv preprint arXiv:1811.07793*, 2018.

[7] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] Suyog Dutt Jain, Bo Xiong, and Kristen Grauman, "Pixel objectness," *arXiv preprint arXiv:1701.05349*, 2017.

[9] Matthew D Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[10] Shih-Syun Lin, I-Cheng Yeh, Chao-Hung Lin, and Tong-Yee Lee, "Patch-based image warping for content-aware retargeting," *IEEE transactions on multimedia*, vol. 15, no. 2, pp. 359–368, 2012.

[11] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[12] Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir, "A comparative study of image retargeting," in *ACM transactions on graphics (TOG)*. ACM, 2010, vol. 29, p. 160.

[13] Philipp Krähenbühl, Manuel Lang, Alexander Hornung, and Markus Gross, "A system for retargeting of streaming video," in *ACM Transactions on Graphics (TOG)*. ACM, 2009, vol. 28, p. 126.