Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-18389-4

Deep learning-based importance map for content-aware media retargeting

Thi-Ngoc-Hanh Le[1] · Tong-Yee Lee[1] · Shih-Syun Lin[2] · Weiming Dong[3]

Received: 6 December 2022 / Revised: 12 October 2023 / Accepted: 19 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

We introduce a deep learning-driven framework for creating an adaptably applicable importance map (A2R-Map) that can be integrated with existing image and video retargeting operators. A conventional retargeting algorithm uses a heuristic approach to seek an off-the-self algorithm used into their retargeting system. The extracted importance map of the image does not match the characteristics of the input image; therefore, it affects the retargeting results and limits the performance of the retargeting method. Our designed framework attempts to minimize the artifacts/distortions caused by inappropriate energy, *e.g.*, the shrunk phenomenon in warping-based results and carving-through-object distortion in the seam carving-based approach. Our proposed framework focuses on capturing sensitive distortion regions and activating their energy to solve this challenge. We verify the effectiveness of our proposed scheme by plugging it in three typical retargeting methods: seam carving-based, warping-based for image, and video retargeting. Extensive experiments and evaluations are conducted on two widely used databases. On the one hand, A2R-Map significantly reduces the time of importance map generation in retargeting systems to $\sim$ 9 times compared to the baseline saliency map. On the other hand, our A2R-Map achieves improvement over the baseline methods with an average of 11% and 9% in terms of image and video quality, respectively. The experimental results and evaluations demonstrate that our strategy for A2R-Map substantially outperforms the previous works and significantly boosts the visual quality of video/image retargeting.

✉ Tong-Yee Lee
tonylee@mail.ncku.edu.tw

Thi-Ngoc-Hanh Le
ngochanh.le1987@gmail.com

Shih-Syun Lin
catchylss@gmail.com

Weiming Dong
weiming.dong@ia.ac.cn

[1] National Cheng-Kung University, Tainan, Taiwan, ROC

[2] National Taiwan Ocean University, Hsinchu, Taiwan, ROC

[3] Institute of Automation, Chinese Academy of Sciences, Beijing, China

Published online: 15 February 2024

Keywords Retargeting · A2R-Map · Seam carving · Warping

1 Introduction

Image and video have long been the widespread media forms in our life. The development of media platforms (e.g., Facebook Reel, TikTok, Instagram, Youtube, etc.) along with the evolution of heterogeneous devices requires the media forms to be well displayed in different resolutions and aspect ratios. This impulse has made media retargeting a more active and attractive research topic in computer vision and computer graphics during the last decade.

This problem has been explored. The conventional content-aware image/video retargeting methods [1–7] rely on the visual information of the image/video to define the importance of the image/video, which should be preserved after retargeting. These methods obtain the content analysis via existing techniques, e.g., saliency map, gradient map, depth map, structure map, shadow map, etc. The result after this analysis is called an "*importance map*", in which an importance value is assigned for each pixel. The important regions of the image must have a higher importance value to be effectively preserved in retargeting process [8]. Usually, a particular method uses a heuristic to seek an approach that could be integrated into their retargeting system. However, these methods are not originally designed for retargeting. The extracted importance map of the image does not match the characteristics of the input image; therefore, it affects the retargeting results and limits the performance of the retargeting method [8]. The latest deep learning methods [9–12] can improve the performance in image retargeting, especially in extracting the importance map of the image. However, it requires equipment with high computing power and comprehensive datasets. Unfortunately, ideal retargeting results are limited and not available. These inadvertently become challenges in this research domain. In addition, each category of retargeting technique has its advantages and limitations. Warping-based methods can produce smoother results without loss of image information, but the shape of the objects of the image is shrunk. Since the seam carving algorithm alone could not perform well, there was a tendency to combine it with other operators such as scaling [8].

In this paper, we propose a framework to address the above challenges. We aim to generate an energy map that could be adaptable to seam carving and warping operators. Our designed framework attempts to minimize the artifacts/distortions caused by inappropriate energy, as we visualize in Fig. 1. The proposed framework pays attention to capturing sensitive-to-distortion regions and activating their energy. Our framework consists of an online learning and an offline refinement stage. The online stage learns the features of the input image to define the region of the main object and predict energy for the pixels belonging to such regions. We achieve this by proposing a neural network model, *Triplet-Layer Features Sharing* (TFS-

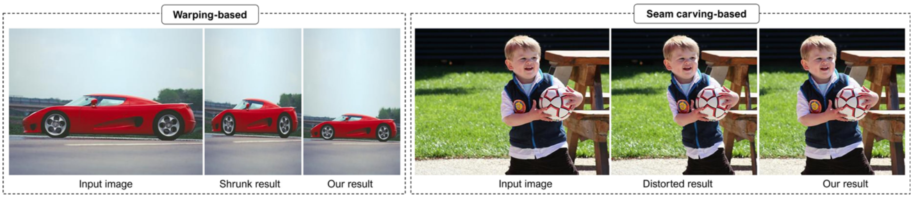


Fig. 1 Our proposed framework resolves the shrinking phenomenon in the warping-based method and carving distortion in the seam carving-based method

Net). The idea of TFS-Net is that we utilize the annotation of salient object detection to explore the important region in an image, which can prevent us from distorting these regions. In the refinement stage, we aim to detect the wrongly predicted energy in the online stage. We activate the important information in the background, along with correcting the energy. This refinement produces a fine and proper energy map, which saves the retargeting results from distortion in the less important regions in the cases where the image content is dense. With this strategy, our energy map enables retargeting methods to face various images. To validate the effectiveness of our scheme, we plug our energy map into the seam carving and warping operators. We test it on image and video retargeting with various input images/videos. More ideal results are obtained. We also compare our results to prior methods in retargeting and importance map generation.

The contributions of our work could be included in the following aspects:

- We develop a framework that effectively defines important maps (A2R-Map) in image and video retargeting applications.
- Our proposed scheme could be adaptable to seam carving and warping-based retargeting systems.
- Our A2R-Map substantially outperforms baseline methods, particularly achieving approximately 11% and 9% improvement over them in terms of image and video retargeting quality, respectively.
- The ideal retargeting results obtained by our system enable researchers in seeking the dataset for this research domain.

We organize the remainder of this paper as follows. In Section 2, we review the prior works that are related to our current research. In Section 3, the detail of our proposed framework is described. In Section 4, our experimental results and evaluations are presented. The conclusion and our future work are discussed in the last section.

2 Related work

The conventional techniques for Content-Aware Image Retargeting (CAIR) are probably categorized into discrete and continuous methods [8, 13]. Most of the resizing systems in the two categories share the mutual process regarding the importance map generation. That is, they all analyze the content of the input image to define the critical regions in advance, which are preserved in the second step of the CAIR procedure. Each CAIR scheme may integrate with a different importance map extraction method. In Table 1, we summarize the techniques that the typical CAIR schemes, including cropping, seam carving, warping, and recent deep learning-based models, use in their framework. For more works, readers are encouraged to refer the survey article [8].

Cropping, a naive technique used in resizing an image, identifies the image's most important content to select the cropping window's location. Researchers in this category define the cropping window in various ways, such as semantic information [1], Support Vector Machine [14], or the gaze of a user looking [15]. For cropping technique, retargeted results are not distorted or damaged the structure. Yet, they can only have one cropping window, in the events that images have several salient and important objects, losing of information outside the cropping window is a negative side of such cropping schemes.

In the seam carving (SC) algorithm, the importance map plays an essential role since the SC seeks to find low-energy seams in the image. The pure SC algorithm [3] defines pixel-energy using the image gradient. Since the gradient map focuses on the object's edge, it leads

Table 1 Overview of existing CAIR methodologies

Methods	References	Technique of importance map
Cropping	Suh et al. [1], Li and Ling [14], Santella et al. [15]	face information [1], SVM [14], gaze of users [15]
Seam Carving	Avidan and Shamir [3], Guo et al. [16], Shen et al. [17], Wu et al. [18], Choi and Kim [19], Battiato et al. [20]	gradient map [3], saliency + gradient map [16], depth map [17, 18], gradient vector flow [20], gradient + saliency + depth + structure maps [19, 21]
Warping	Zhang et al. [22], Guo et al. [23], Wang et al. [24], Zhang et al. [25], Jin et al. [26], Niu et al. [27], Lin et al. [6], Hu et al. [28], Panozzo et al. [29], Tan et al. [30], Kim et al. [31], Kim et al. [32]	distortion map [22], human body extraction [23], saliency map [6, 23–32]
Deep learning approaches	Liu et al. [33], Guo et al. [34], Song et al. [35], Lin et al. [12], Wang et al. [36], Tan et al. [10], Ahmadi et al. [37], Cho et al. [9], Zhou et al. [38]	Convolution Neural Network

to distortion passing through the objects. Inspired by this, several works [16–19, 21, 39, 40] subsequently investigate various ways to address the drawback of the gradient map. They could be saliency map, depth map, structure map, or combine these maps, as outlined in Table 1. Although these solutions help improve SC's performance comparing to the baseline [3], they still have significant downfalls. The images with background color is close to the color of important regions, foreground with multiple objects, or dense of background content are challenging to them.

CAIR methods in the category of the warping-based attempt to minimize the deformation of regions of high visual importance, while higher deformation is allowed in regions of low importance [8]. Hence, a proper content analysis method integrated into such a warping scheme also plays an vital manner. Each work in the warping-based approach utilizes a different way to construct the importance map of the image. Along with distortion map used in the system of Zhang et al. [22], the saliency map is the most used technique, which is used in most of the warping schemes [6, 23–32]. Nonetheless, these importance map generation techniques are not designed for retargeting application. This leads to linear changes in the shape of resizing results, which are the common drawbacks in these warping-based systems.

Recently, deep learning-based technologies are investigated to explore retargeting domain [9, 10, 12, 33–38]. These state-of-the-art approaches focus on two sides. They utilize Convolutional Neural Network (CNN) to define the importance map and then feed this resulting map to a retargeting operator [35, 37]. On the other side, they develop a model based on the existing retargeting concept, e.g., warping [9, 10], seam carving [12], and multi-operator [38]. The advances in deep-learning techniques are now able to boost the research of analyzing multimedia content [41]. Several applications benefit from this evolution, for example, classification [42], and object detection [43]. Inspired by these and the observation of the drawbacks mentioned above in image/video retargeting applications, in this work, we take advantage of the deep-learning technique to boost the performance of the existing retargeting methods. In contrast to prior work in the retargeting domain, our approach is to generate an energy map that could be adaptable for a particular retargeting method. We consider both background and foreground information in the importance map generation, which is efficient in avoiding carving distortion in the seam carving-based approach and shrinking phenomenon in the warping-based system.

3 Methodology

3.1 System overview

Our proposed framework is illustrated in Fig. 2, which consists of an online and offline stages. The system gets as input a color image $\mathcal{I}$ and we aim to generate the corresponding Adapt-to-Retarget importance map (A2R-Map). We also call A2R-Map in the term "energy map" in our article. The online stage is used to estimate the energy of the pixels that belong to the most important region in $\mathcal{I}$. For this stage, we propose a network called Triplet-Layer Feature Sharing (TFS-Net). We train TFS-Net to automatically produce an energy map, denoted as OMap. The offline stage is a refinement manner. We first extract edge features in the image $\mathcal{I}$ to obtain an energy map, denoted as BMap. Thereafter, we formulate OMap and BMap to define the final A2R-Map.

3.2 OMap generation

We design the TFS-Net to shoulder the task of OMap generation. Our TFS-Net is configured with two modules, a feature extractor and a feature sharing session, as illustrated in Fig. 2. As named, the first module is to extract the features in the input image. For this purpose, we use VGG-19 [44] as a backbone. The pre-trained VGG-19 is widely used as the backbone network in many applications, particularly for silent object detection (SOD) tasks. Therefore, it is reliable to be considered a good feature extractor. Furthermore, VGG-19 has been trained on the large-scale dataset. With this strategy, we can remove the burden of training for this process. It's worth noting that VGG-19 is originally designed for image classification, which is structured by feature extraction and classification parts. To use it as a backbone, we remove

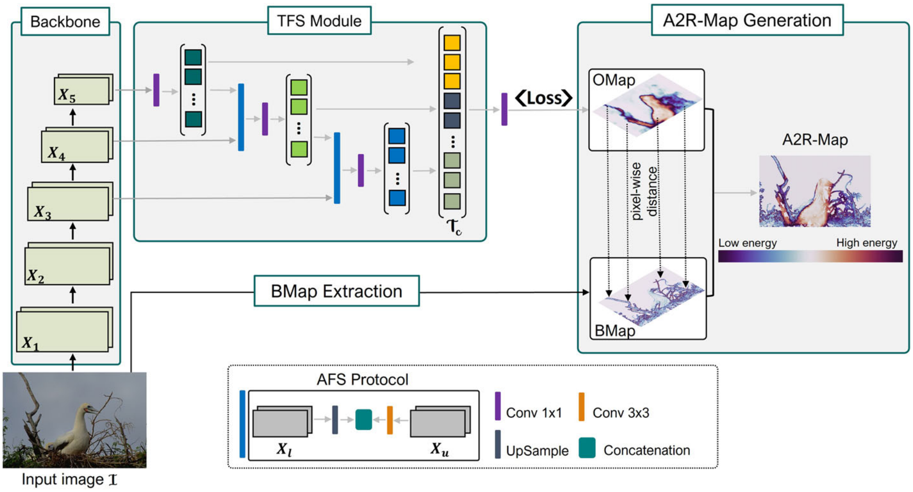


Fig. 2 Our proposed framework of A2R-Map generation

the second part and only use the first part for feature extraction purposes. In the second module, the so-called feature sharing session (TFS), we propose to learn feature correlations extracted from the backbone and formulate them to estimate pixel energy. We note here that the feature extraction of VGG-19 is designed with 5 layers, structured as *Convolution* + *ReLu* → *Max-pooling*, and finalized by 3 layers of *Fully connected* + *Relu*. The last 3 layers are used to link the extracted features to the output, rather not for extracting feature purpose. Hence, we only use the first 5 layers in this backbone. Given an input color image $\mathcal{I}$ in size of $H \times W \times 3$, where H and W are the height and width, we can obtain five-layer features $\{\mathcal{X}_i | i = 1, \ldots, 5\}$ with sizes $[\frac{H}{2^i}, \frac{W}{2^i}]$ from the backbone network. Feature $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_3$ with larger sizes are low-level features with rich object information, $\mathcal{X}_4$ and $\mathcal{X}_5$ are high-level features with rich semantic information. Besides, feature $\mathcal{X}_1$ brings much computation cost and slight performance improvement. Therefore, to take the advantage of the features of the layers, we use the last three-layer features for subsequent processing.

Once the image $\mathcal{I}$ is encoded by the backbone, a triplet of the last three-layer features $(\mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5)$ is fed to the TFS session. We do not use the features in the first two layers since they are relatively coarse. The deeper layers capture more high-level features, which are more beneficial for the resultant energy map in challenging complex image content. The TFS shoulders the task of learning the feature representation of the input $\mathcal{I}$, finding their correlation to predict the proper energy of the pixels belonging to the important region of $\mathcal{I}$. To achieve this, we embed into TFS an Adjacent-layer Features Sharing (AFS) protocol. To be more specific, AFS is designed to let any two adjacent layers share their features. Features of each layer are first convoluted with a distinct number of filters. They are then concatenated to yield the product feature maps, denoted by $\mathcal{S}_{l \to u}$. Mathematically, this process is formulated as:

$$\mathcal{S}_{l \to u} = \mathcal{F}\Big(C^3\big(\zeta(\mathcal{X}_l), k\big), C^3(\mathcal{X}_u, r)\Big), \tag{1}$$

where $\mathcal{F}$ is the concatenation; C^3 is operated by a convolution with the kernel size of 3×3 operator. We use a 3×3 kernel since it has a smaller receptive field compared to larger kernels. This means it focuses on capturing more local features, which can be helpful for detecting fine details and edges in the image. Plus, it requires less computation cost. k and r denote the number of filters of lower $\mathcal{X}_l$ and upper $\mathcal{X}_u$ layers, respectively. And ζ indicates the up-sample operator.

As we illustrate in Fig. 2, applying AFS protocol on each pair of adjacent layers yields a product feature maps $\mathcal{S}_{i \to i-1}$. The sharing product at layer i is recursively defined as:

$$\mathcal{S}_{i \to i-1} = \begin{cases} \gamma\big(\mathcal{S}_{i+1 \to i}, \mathcal{X}_{i-1}\big) \text{ if } i < 5 \\ C^1(\mathcal{X}_i) \text{ if } i = 5 \end{cases}, \tag{2}$$

where $\gamma(.)$ is the AFS protocol expressed in (1); C^1 is a convolution 1×1. Thereafter, the products defined by (2) are fused together to construct the final tensor $\mathcal{T}_c$. We finally pass $\mathcal{T}_c$ through a 1×1 convolution to map it into the ground truth with an activation function. In the training process, we use a sigmoid activation function to calculate the probability as an output that has a value in the range of 0 and 1. All the parameters in our network are learned by minimizing the loss function, which is computed by the errors between the probability map and ground truth. Given a ground truth $S_g (S_g \in 0, 1^{h \times w})$, which is corresponding to the input image $\mathcal{I}$ $(H \times W \times C)$, stochastic gradient descent is employed to minimize the loss of training to predict visual information probability:

$$\mathcal{L}(\mathbf{S}_g, \mathbf{S}_p) = -y_i \times log(\hat{y}_i) + (1 - y_i) \times log(1 - \hat{y}_i), \tag{3}$$

where $\mathbf{S}_p$ is the estimated energy map produced during the training; $y_i \in \mathbf{S}_g$ and $\hat{y}_i \in \mathbf{S}_p$. After training TFS-Net with the loss function (3), pixels in the input image $\mathcal{I}$ are predicted as important degree by being assigned an energy value ranging from $[0, \ldots, 1]$. The higher value indicates the pixel belongs to such an important region. We call this estimated result an OMap, which is then further formulated in the following step to define the final importance map.

3.3 A2R-map generation

The resultant energy map obtained by the TFS-Net, *i.e.,* OMap, is sufficient to improve the performance of seam carving-based and warping-based methods in the game of retargeting. This effectiveness is discussed by the ablated results in Section 4.2. Nevertheless, retargeting is a particular application in which a proper definition of pixel-wise energy plays an essential role [8]. Since the benchmark dataset of such an energy map is not available, we utilize the annotation data of salient-object-detection to train our TFS-Net. As a result, TFS-Net focuses on the region of the labeled objects and may skip the objects belonging to the background. Such a resultant energy map could be good for images that are with simple content, *i.e.*, one object in the foreground and the background is not complex. In the cases that the input images are with dense backgrounds, this may lead to loss the semantics of the retargeted image, *i.e.*, some significant pixels are invisible and distorting artifacts occur at these regions. To overcome these challenges, we further do a refinement based on the initial energy obtained from TFS-Net (*i.e.*, the OMap). The pseudo-code of this strategy is presented in Algorithm 1.

Algorithm 1 Algorithm of A2R-Map generation.

```
Input: Color image $\mathcal{I}$
Output: Energy map A2R-Map
Train TFS-Net
$OMap \leftarrow$ TFS-Net($\mathcal{I}$)
BMap $\leftarrow \sqrt{g_x^2(B_x, \mathcal{I}) + g_y^2(B_y, \mathcal{I})}$
Put the grid $8 \times 8$ on $OMap$ and $BMap$
$\mathcal{W} = \{NULL\}$
for each pair $(w_i^{OMap}, w_i^{BMap})$ do
    $\Phi_i \leftarrow \sqrt{\sum_{k=1}^{n} |w_i^B - w_i^O|}$
    if $\Phi_i > \eta$ then
        Add $w_i^{OMap}$ to $\mathcal{W}$
    end if
end for
for each $w_i \in \mathcal{W}$ do
    if $d\big(p^O(i,j), p^B(i,j)\big) > 0$ then
        A2R-Map$(i,j) =$ BMap$(i,j)$
        if $d\big(p^O(i,j), p^B(i,j)\big) < 0$ then
            A2R-Map$(i,j) =$ OMap$(i,j)$
        end if
    end if
end for
Return the energy map A2R-Map.
```

To begin, we make the edge features of input image $\mathcal{I}$ to be visible. An energy map of this manner is accordingly produced, denoted as BMap, which can be formulated as:

$$\text{BMap} = \sqrt{g_x^2\big(B_x, \mathcal{I}\big) + g_y^2\big(B_y, \mathcal{I}\big)}, \tag{4}$$

where B_x, B_y are the Sobel kernel [45] of horizontal and vertical, respectively. And g_x, g_y are the two images which at each point contain the horizontal and vertical derivative approximations respectively. Thereafter, we base on BMap to adjust the energy in OMap. The reason is that, with (4), we make the edge pixels visible and they are assigned high energy values. Hence, we can treat BMap as a standard map to allocate the "*object*" boundary. "Object" here includes both the main objects in the foreground and the objects in the background. Our adjustment on OMap focuses on two aspects: (1) detecting the wrongly estimated energy in OMap, and (2) activating the important information in the background that could be missed in OMap. It is worth pointing out that this refining strategy is different from combining the two maps. Combining leads to the wrongly predicted energy in OMap still exists. This eventually affects the content structure of the retargeting results.

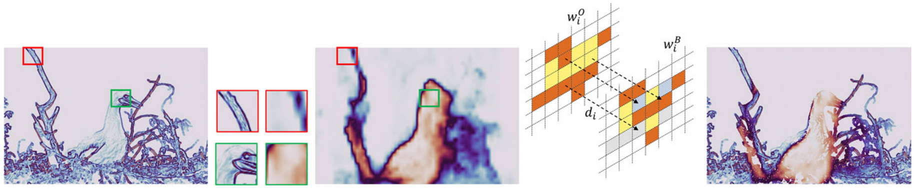
Given two energy images OMap and Bmap corresponding to the input image $\mathcal{I}$, we simultaneously slide a window w on the two maps. The size of w is set to 8 × 8. A smaller w leads to higher computation cost, and content in small w not is sufficient to define the inconsistency. Meanwhile, a larger w spends less cost, but reduces the accuracy of BMap. We primarily test on different sizes of w and conclude that w-size in range of 8 to 12 guarantees performance of BMap to be stable with arbitrary image content. All of experiments in this article, we use w of 8 × 8. Let us denote w_i^O and w_i^B as the window capturing OMap and BMap at the iteration i. We then calculate the distance of pairwise windows as:

$$\Phi_i = \sqrt{\sum_{k=1}^{n} \left| w_i^B - w_i^O \right|}, \tag{5}$$

with n is the total pixel in w_i. A large Φ_i reveals the "*inconsistency*" between w_i^B and w_i^O. "Inconsistency" here refers to the wrongly predicted energy of pixels in w_i^O. For example, we can see the visualization of this phenomenon in Fig. 3, highlighted in green and red squares. It is observed that there is significant difference between them. At first glance, OMap seems to be good. However, zoom-in each window shows a significant inconsistency in terms of the spatial location of the input content and predicted energy. This phenomenon eventually affects the structure of retargeting results.

We define the set of patches that encompass of the wrongly-predicted-energy pixels as:

$$\mathcal{W} = \{w_k \in \text{OMap } s.t.\ \Phi_k > \eta\}, \tag{6}$$



(a) BMap (b) Pair of windows (c) OMap (d) Pixel-wise distance (e) A2R-Map

Fig. 3 Samples of inconsistent pairs

with η is the threshold we set in our experiment, *i.e.*, $\eta = 50$. We note here that the threshold η linearly varies with the size of window w. As we mentioned above on the stable range of w, the range of η is recommended in range of $[50, \ldots, 55]$. Once $\mathcal{W}$ is found, we define $A2R$-Map as:

$$\text{A2R-Map}(i, j) = \begin{cases} \text{BMap}(i, j) \text{ if } d(p^O(i, j), p^B(i, j)) > 0 \\ \text{OMap}(i, j) \text{ if } d(p^O(i, j), p^B(i, j)) < 0 \end{cases}, \tag{7}$$

here p^O, p^B are the pixels belonging to OMap and BMap, respectively. We note here that the pixel-wise distance $d(p^O, p^B)$ is used in this equation to define which pixel could be used to update the corresponding pixel in A2R-Map.

Our above refinement phase encourages the predicted energy in OMap to be consistent with the content in the input $\mathcal{I}$, see Fig. 3(e). Besides, it makes the essential information in the background to be visible. Therefore, our A2R-Map can facilitate seam carving from carving wrong energy and saving the warping results from the shrinking effect.

4 Experimental results

In this section, we first present our experimental settings. In Section 4.2, we analyze how our A2R-Map affects the image and video retargeting performances and discussion via the ablated results. Finally, we show the visual comparisons and apply some evaluating indicators to evaluate the performance.

4.1 Implementation details

We implemented our system on the PC with Intel Core i7 CPU, 16GB RAM, and NVIDIA GeForce GTX1070 GPU. The language is used in our importance map generation is Python 3.6. To generate retargeting results, the seam carving operator is implemented in Python, and warping-based is computed with C++ programming language in Visual Studio 2015. In our TFS-Net network, we use MSRA 10K dataset [46], which is used for saliency detection, as our training data. This dataset consists of 10000 images with a diversity of the content structure of natural scenes. The dataset also contains manually annotated ground-truth saliency. In terms of parameters settings for each approach used in our comparisons and evaluations, we summarize in Table 2.

Table 2 Description of parameters in compared methods

Method	Description
Baseline warping for image [6]	Importance map: Saliency [47] + segmentation [48]
Baseline warping for video [7]	Importance map: Saliency [47] + segmentation [48]
Baseline SC [3]	Importance map: gradient energy
Patel et al. [49]	The same parameter as the source paper [49]
RC map [50]	The same parameter as source paper [50]
NIF [16]	The same parameter as source paper [16]
BASNet [51], DIS [52], DFI [53]	The same parameter as source papers

4.2 Image and video retargeting with A2R-map

Given an image/video $\mathcal{A}^s$, arbitrary image/video retargeting methods $\mathcal{R}$ aim to generate a target image/video $\mathcal{A}^t$ with the following function:

$$\mathcal{A}^t = \mathcal{R}\big(\mathcal{M}(\mathcal{A}^s), \mathcal{P}\big), \tag{8}$$

where $\mathcal{P}$ denotes the resizing operator of $\mathcal{R}$; $\mathcal{M}$ is an off-the-shelf method that $\mathcal{R}$ uses to define the importance in the input $\mathcal{A}^s$. As we discuss in the aforementioned section, we aim to generate an energy map that could be adaptable to an arbitrary resizing method $\mathcal{R}$ and eliminate the artifacts caused by inappropriate energy of $\mathcal{M}$. In other words, the method of importance map generation $\mathcal{M}$ is alternated by our A2R-Map in the methods $\mathcal{R}$. In this section, we verify the effectiveness of our proposed A2R-Map by plugging it into three typical retargeting methods: seam carving-based, warping-based for image, and video retargeting.

Seam carving-based We compare the ablated results when integrating seam carving operator with gradient map, OMap, and A2R-Map in Fig. 4. As shown in this figure, integrating the seam carving operator with different energy maps yields different results. Using gradient maps results in deformed salient objects since these maps can only indicate high energy near the edges of an object [16]. The result in (b) visualizes such distortion. In (c), we can see that the OMap obtained by our TFS-Net demonstrates its benefit in overcoming the mentioned phenomenon in (b). However, the wrongly predicted energy in the background pixels causes obvious carving artifacts in the background. Thanks to the refinement of our approach, our A2R-Map resolves these phenomena. The result in (d) reveals that A2R-Map serves a better result compared with the two shown cases.

Warping-based Here we demonstrate the effectiveness of our A2R-Map in the warping-based systems for image and video retargeting. We adopt [6] and [7] as the case study for image and video retargeting, respectively. These two works are mentioned as good warping schemes. Besides, the source codes are provided by the authors, thus they are reliable to use and fair for comparisons. The problem in warping-based retargeting results is different from those in the seam carving-based approaches, i.e., the results are shrunk. The reason is that these warping schemes rely on the saliency map, adopted from Goferman et al. [47], to estimate the moving factor of quad vertices. That is, the vertices with high saliency value are assigned a small moving weight. And vice versa, they are assigned the same scaling weight. As a result, in the case that the saliency value is not correct, the quad vertices tend to scale linearly.

Figure 5 visualizes this phenomenon, and we further use the linear scale's results in these comparisons. For the image retargeting result (*i.e.*, the first row), by observing the three

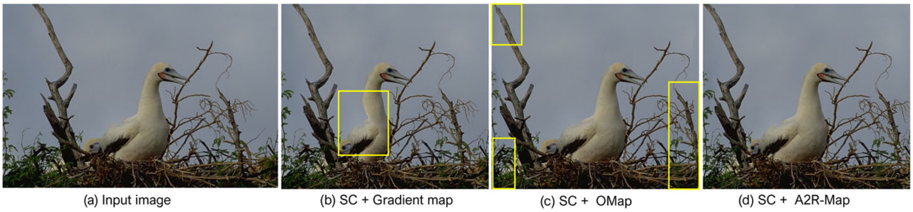


Fig. 4 Results of seam carving operator using different energy maps

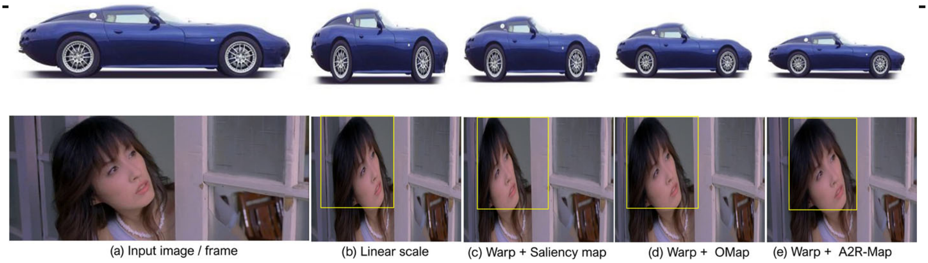


Fig. 5 A2R-Map with warping-based methods on image and video retargeting

results (c)-(e), we can see different energy maps yield different retargeting. At first glance, the saliency map [47] could be a good partner to integrate with warping [6, 7]. However, the shape of the car is not preserved well. This effect makes Lin's results relatively close to the linear scale. As in our prior discussion, OMap alone is sufficient to resolve the shrinking issue in such a warping-based system. The result in (d) demonstrates this effectiveness. Nevertheless, A2R-Map boosts the result more ideal, *i.e.*, the shape of the car is quite similar to those in the input image. It's obvious to see that the result in (e) outperforms the ablated results in (c) and (d). The second row further demonstrates the benefit of our A2R-Map in terms of video retargeting. We can observe differences in the region highlighted in yellow. The shape of the girl's head is distorted significantly. Similar to the shown case of image retargeting, OMap improves the problem occurring in (c) but is not good as A2R-Map's performance. The visualization for videos can be seen at this link[1].

4.3 Our results and discussion

Here we give out more discussion on the capability of our A2R-Map. We test the images with low retargetability. The data is obtained from [54]. We examine on two samples shown in Fig. 6, one is with medium degree (*i.e.*, retargetability: 0.57) and one is with a low score (*i.e.*, retargetability: 0.1). Figure 6 shows the plausible results generated by our system in this manner. As shown in the figure, the shape of the main object (*i.e.*, the clock) is distorted by deformation in AAD [29] (Fig. 6(a)), and the important regions are cropped (*i.e.*, the hand and the mug in Fig. 6(b)). In contrast, these phenomena do not occur in our results. These experiments imply that our method is tolerated low retargetability images. Furthermore, the images with reflection symmetry are challenging when retargeted by seam carving operator [49]. The authors in [49] propose a novel method to address this problem. Figure 7 visualizes the results when our A2R-Map competes with [49] on a sample containing reflection symmetric objects. The yellow rectangles highlight the differences between results. In this input image, the head of the zebra contains reflection symmetric attributes. In this regard, [49] and our A2R-Map are successful in preserving such objects and are quite better than the gradient map. However, other regions (e.g., the leg of the zebra or the background on the top-left corner) are distorted in [49]'s result. The comparisons in Figs. 6 and 7 reveal that our approach is effective with various challenging input images. This enables the existing retargeting method to have more ideal results. Appealing results are also produced. Readers can explore our project website[2] for more experimental results, including image and video retargeting.

[1] http://graphics.csie.ncku.edu.tw/A2RMap/CompareVids.mp4

[2] http://graphics.csie.ncku.edu.tw/A2RMap

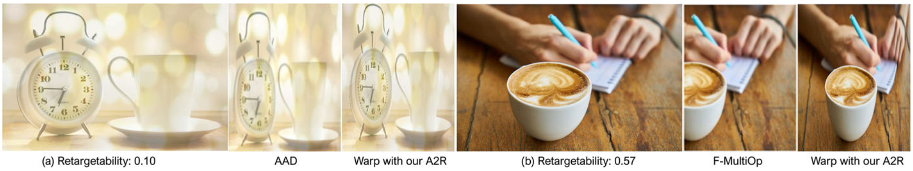


Fig. 6 Our A2R-Map challenges on the low retargetability images

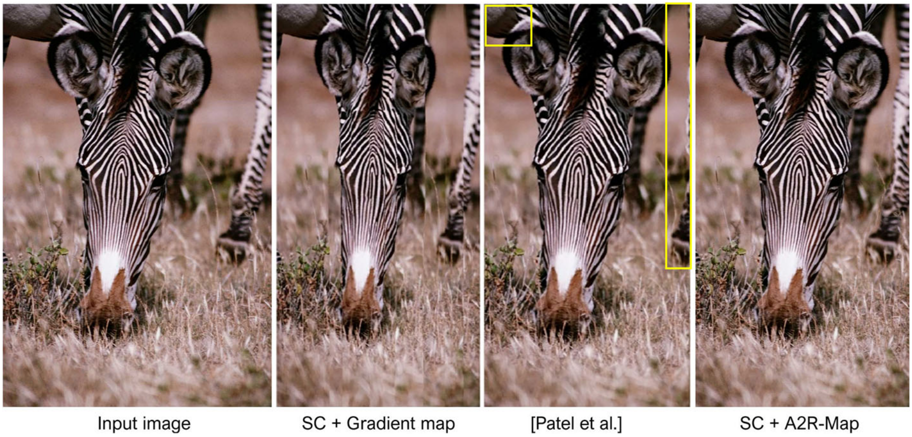


Fig. 7 Comparison on symmetry image with Patel et al. [49]

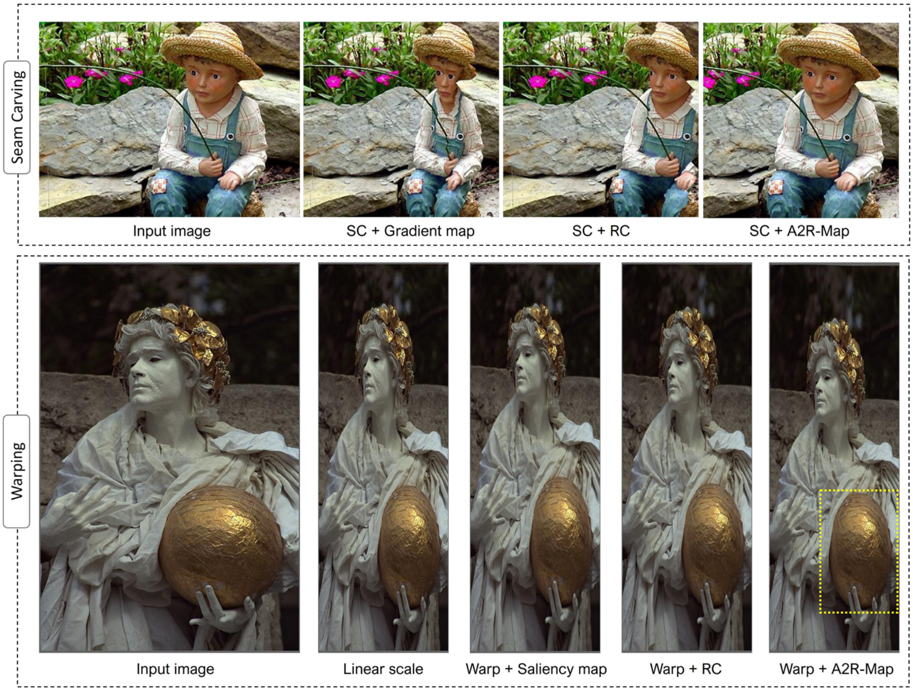


Fig. 8 Our A2R-Map and RC map [50] on seam carving and warping operator

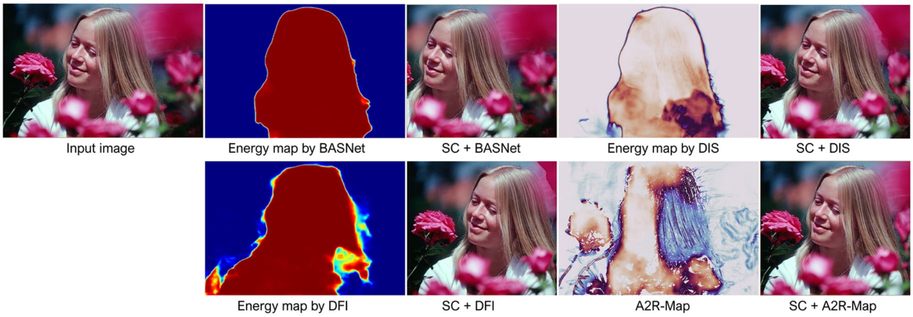


Fig. 9 Our A2R-Map competes with SOTA SOD models

It's worth hypothesizing that $\mathcal{M}$ in (8) is another saliency detection method. That is, instead of plugging our A2R-Map in such video/image retargeting system $\mathcal{R}$, the method $\mathcal{R}$ can choose other $\mathcal{M}$. To justify this hypothesis, we select four candidates, RC [50] and three SOTA SOD models BASNet [51], DFI [53] and DIS [52]. RC is a global contrast-based saliency region detector introduced by Cheng et al. [50]. This method has been mentioned as an efficient importance information generator in such a retargeting system [32]. Three opted models are good salient object detection methods used to detect salient objects in many applications. We plug these four methods in (8) and compare their retargeting results against our A2R-Map. The visual results are exhibited in Figs. 8 and 9. The results demonstrate that our A2R-Map outperforms in all cases. RC seems unsuitable for seam carving operators since their result has considerable carving distortion. For warping operators, RC shows its adaptable capability. However, RC still suffers the drawback of "*linear-like*" as we discussed above. In terms of SOTA SOD models in Fig. 9, since these SOD methods are originally designed for object detection, it makes sense to find that non-labeled-objects (*i.e.*, the flowers) are carved.

4.4 Visual comparisons

To demonstrate that our approach advances prior work in retargeting, we compare it with five methods. For a thorough and fair comparison, we divide this session into two groups, seam carving-based approach and warping-based one. In the first group, three methods [16, 21, 35] are compared. The mutual point in these methods is that they attempt to produce an importance map that could resolve the distortion in the seam carving operator. They approach this problem differently, combining different image features [16, 21] or modeling a neural network [35]. In terms of warping, two deep learning-based techniques investigated in recent years [9, 10] are compared.

Figures 13, 12 and 14 demonstrate the comparisons in terms of seam carving operator. NIF [16] considers texture information together with color information to construct an effective energy map. To achieve this, Guo et al. [16] combine the gradient map and a saliency map. However, in the cases that the background is more inhormogenous than the important areas, their algorithm fails. Figure 13 visualizes such cases and their results. Also shown in this figure, our energy map A2R-Map estimates the importance map better than NIF's. As a result, our retargeting result is quite better NIF's result and without any distortion. In Fig. 12, although [21] combine various image features to define the energy map for seam carving operator, their performance in this example is not good as ours. For example, they perform well in the region of the butterfly, the left and right side of the image, but some noticeable distortion on the body of the tree makes their retargeted result is not as ideal as ours. CarvingNet [35] is mentioned as the earliest work that investigates a neural network to generate the importance map to improve seam carving-based method. Figure 14 is the visual comparison between our results and theirs. In this challenging case, *i.e.*, the content in the input image is quite complex, CarvingNet creates obvious distortion at the building and the tree (highlighted in yellow rectangles). Meanwhile, our A2R-Map performs better, *i.e.*, there does not exist noticeable distortion and the shape of the heart-shaped balloon is not significantly scaled down as in CarvingNet.

Figures 15 and 16 visualize the comparisons in terms of warping-based methods. As shown in the results, both WSSDCNN [9] and Cycle-IR [10] share the same drawback of preserving the structure of the input image in the warping session of their network. Their results are good in the structure of the bird or the house, but the background contents are damaged (see highlighted region in both cases). However, our results are quite better in this competition. For more comparisons, readers are encouraged to explore the Appendix and our project website.

4.5 Objective evaluation

To quantitatively evaluate our method's performance, we first adopt two metrics, ARS [55] and Sift-Flow [56]. ARS algorithm is a metric that evaluates the visual quality of retargeted images by exploiting the local block changes with a visual importance pooling strategy. We use this metric to evaluate the distortion degree of our results comparing to the corresponding input image. For the Sift-flow, we use it to estimate dense correspondence between the original and retargeted images in term of image content preservation. In both metrics, the higher is better. In our evaluation session, we use RetargetMe database [57] as the benchmark data. We compare the two metrics on the results generated by five methods: SC + Gradient map, SC + A2R-Map, Warp + Saliency map [47], Warp + A2R-Map, and Cycle-IR [10]. We note here that the abbreviations SC and Warp refer to seam carving operator [3] and warping operator [6]. The visual results of this session can be found in our project website. The analysis results on two metrics are presented in Table 3. The analysis results reveal that our A2R-Map can generate results with less distortion than other competitors. This advantage is demonstrated by the higher ARS score, approximately 11% on average. In term of Sift-flow score, Cycle-IR is higher than ours when our A2R-Map integrates with SC; but A2R-Map+Warp has a relatively comparable effectiveness with Cycle-IR in this manner.

A part from above metrics, we further adopt the bidirectional similarity measure (BSM) [58] to evaluate the quality of retargeted images and videos. Simakov et al. [58] investigate BSM to describe the coherence and completeness between input and output images. It is widely used for quantitative analysis retargeting results in several works. For this metric, we conduct evaluations on two groups. On the first group, we use seam carving operator integating with different ways for energy map generation. Three SOTA models BAS [51], DFI [53], DIS [52], and our A2R-Map join in this competition. Given a pair of images

Table 3 Retargeting quality analysis

	Warp		A2R+Warp		SC		A2R+SC		Cycle-IR	
Metrics	ARS	Sift-flow	ARS	Sift-flow	ARS	Sift-flow	ARS	Sift-flow	ARS	Sift-flow
Avg.	0.82	0.64	0.92	0.69	0.77	0.61	0.85	0.72	0.89	0.79

$(\mathcal{A}^s, \mathcal{A}^t)$, the source image $\mathcal{A}^s$ and the corresponding resized one $\mathcal{A}^t$, the error of $\mathcal{A}^t$ over $\mathcal{A}^s$ is expressed as:

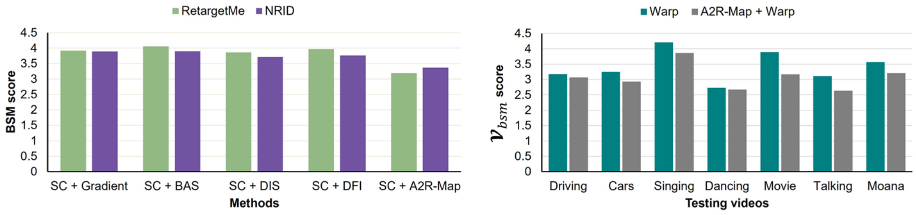
$$\mathrm{BSM}(\mathcal{A}^s, \mathcal{A}^t) = \frac{1}{N}\left(\sum_{p\subset\mathcal{A}^t}\min_{q\subset\mathcal{A}^t}\delta(p,q) + \sum_{q\subset\mathcal{A}^s}\min_{p\subset\mathcal{A}^s}\delta(q,p)\right), \tag{9}$$

where N is the total patches on $\mathcal{A}^s$ and $\mathcal{A}^t$; $\delta(.)$ is defined by sum of squared distance of two patches p and q. The lower BSM represents better retargeting quality. In this group, we examine on two benchmark datasets, RetargetMe [57] and NRID [59], which consists of 80 and 35 images, respectively. The analysis results are presented in Fig. 10-(a). We can see that using the importance map generated by SOD models, *e.g.*, BAS, DIS, and DFI, yields relatively identical effect with gradient-based energy via the light differences in score. Yet, SC combines with our A2R-Map serves better performance with lower BSM scores. All of our competitors perform better on NRID dataset than on RetargetMe, this is contrast to ours. However, our scores on two datasets lower than compared models. This analysis reveals that using such an SOD model to define pixel energy in seam carving-based systems is challenging. Averaging on two datasets, using our A2R-Map improves approximately 15% comparing the usage of the alternatives in this analysis.

On the second group, we evaluate quality of retargeted videos. As the ground truth for video retargeting is not available, we elaborate as follows. Given a video with n frames, we have two sets: a set of the source video frames $\mathbf{S}^s = \{\mathcal{A}^s_i, \ldots, \mathcal{A}^s_n\}$ and the other is those in retargeted form $\mathbf{S}^t = \{\mathcal{A}^t_i, \ldots, \mathcal{A}^t_n\}$. For each pair of frames $(\mathcal{A}^s_i, \mathcal{A}^t_i)$, we apply (9) to define the error of frame $\mathcal{A}^t_i$ over frame $\mathcal{A}^s_i$. Afterwards, we measure the error degree of a retargeted video as:

$$\mathcal{V}_{bsm} = \frac{1}{n}\sum_{i=1}^{n} BSM(\mathcal{A}^s_i, \mathcal{A}^t_i). \tag{10}$$

In this evaluation, we conducted on 9 videos (exhibited on our project website) that have diverse content, *e.g.*, single-moving object, multiple-moving objects, complex backgrounds, or important content distributed in the entire frame. Figure 10-(b) presents the analysis result. It's can be seen that our A2R-Map boosts the quality of videos better than the conventional warping system in all of 9 videos. The scores in videos "*Driving*" and "*Dancing*" are relatively close to the compared method. However, the average score of our opponent is 3.42, meanwhile ours is 3.078 which is approximately 9% improvement of video retaregting quality.



(a)

(b)

Fig. 10 Analysis on BSM metric on image (a) and video (b)

Table 4 Analysis on processing time on different resolutions of input images/video frames (Unit: second)

Resolutions	Saliency map [47]	Gradient map	Our OMap	A2R-Map
1024 × 813	40.17	1.25	0.72	3.49
720 × 480	23.52	0.86	0.69	1.52

4.6 Timing analysis

To analyze the processing time, we conducted all experiments on a PC with Intel Core i7 2.5GHz, 16GB RAM. The comparison on energy map generation process is presented in Table 4. The saliency map [47] is implemented by Matlab, the gradient map, our OMap and A2R-Map are implemented by Python 3.6. As reported, saliency map generation [47] takes a huge computation time. Gradient map is faster than ours. However, the offset between our timing and gradient map is not significant. This is a trace-off between the processing time and the better quality results.

4.7 Limitations

Although our proposed A2R-Map substantially minimizes the distortions in prior retargeting work, it is still not good in some cases when playing with the seam carving operator. We show an example in Fig. 11. It is because of lacking the dataset we use to train, and the OMap is not good in such cases. As a result, our calculation in the refinement manner is not efficient. Yet, we can see that the results are still plausible with the warping operator.

5 Conclusions

This paper introduces a learning-based framework for importance map generation that is particularly useful in image and video retargeting applications. The core contribution of our work is (1) the effectiveness in minimizing the distortion in seam carving operator and shrinking phenomenon in mesh-based warping systems, and (2) enabling the existing resizing operators

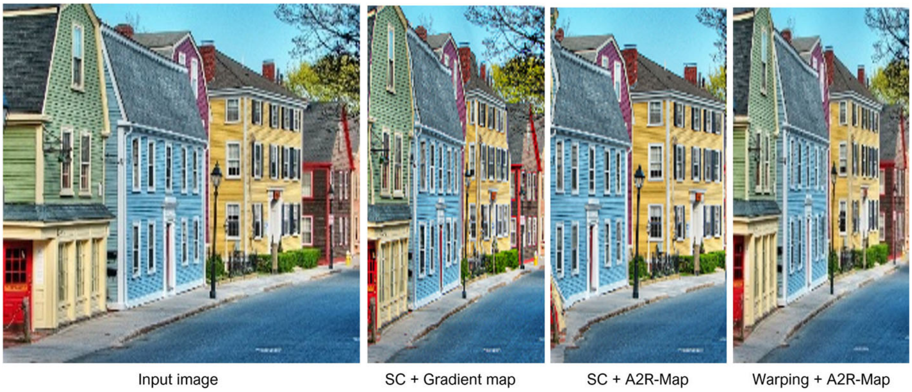


Fig. 11 Our limitation

to challenge various input content images. Our results and comparisons show that the proposed approach substantially outperforms related methods. Furthermore, the experimental results on the low retargetability images and challenging cases are the evidence that reveals the effectiveness of our scheme in retargeting. In our future work, we plan to improve the dataset to alleviate the limitations of this study. Furthermore, given A2R-Map's impressive capabilities in analyzing image and video content, there is potential for us to expand its usage into exploring a novel image and video retargeting system. That is, utilizing A2R-Map to analyze image/video content, then integrating with a diffusion-based technique for resizing manner.

Appendix A: More comparisons

Apart from the comparisons with content-aware retargeting approaches, we further exhibit our results competing with a semantic-aware retargeting approach [60] in Fig. 17. In this figure, besides [60] (PM), we further show the results from other five retargeting methods: seam carving (SC) [3] and its improved version (ISC) [61], patch-based warping (PW) [6], saliency-based mesh parametrization (SMP) [23], multi-operator (MOR) [4]. These results are obtained from [60]. We can observe that our result outperforms the compared results. If the carving distortions occur in SC, ISC, and SMP, a linear-like phenomenon falls in PW, MOR, and PM (*e.g.*, the green door). Meanwhile, our result does not have such phenomena and appears in a balanced structure compared to the input image. Figure 18 exhibits the performance of our A2R-Map in terms of enlarging. In this experiment, we enlarge images to 25% of width.

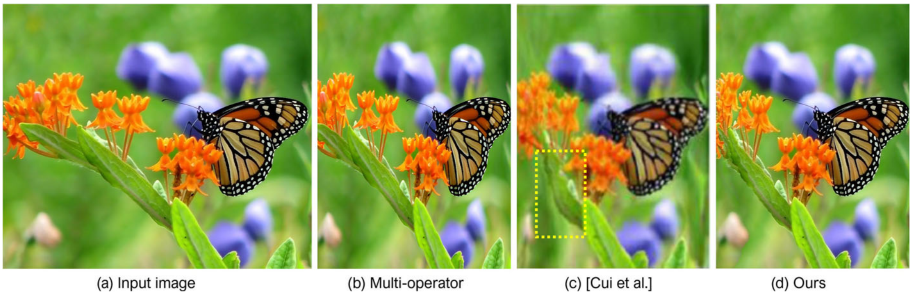


Fig. 12 Comparison with Multi-operator and Cui et al. [21]

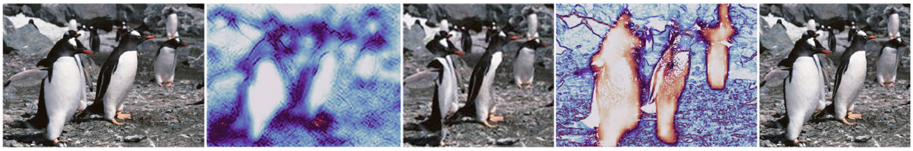


Fig. 13 Left to right: input image, NIF energy map, SC + NIF, A2R-Map, SC + A2R-Map

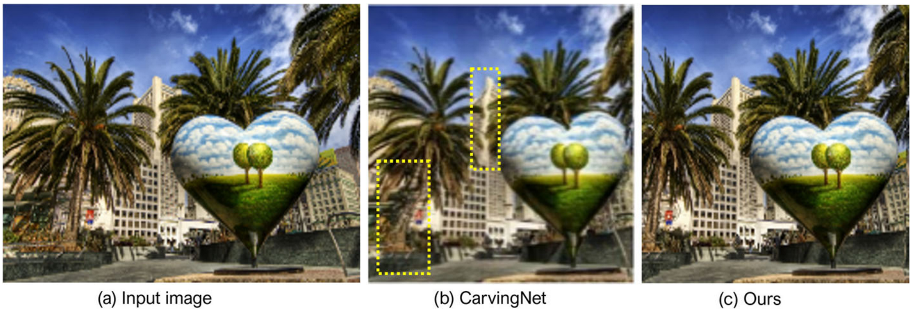


Fig. 14 Comparison with CarvingNet

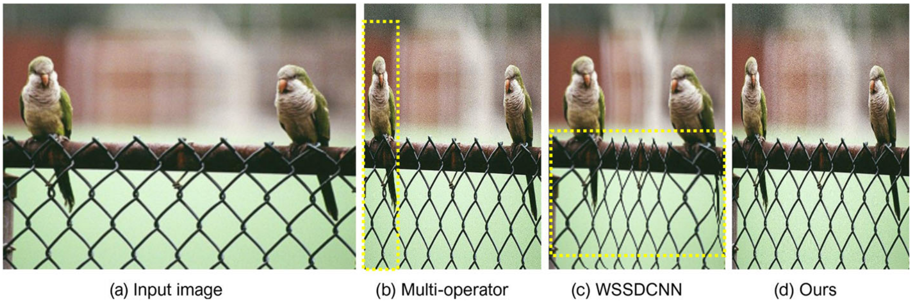


Fig. 15 Comparison with WSSDCNN

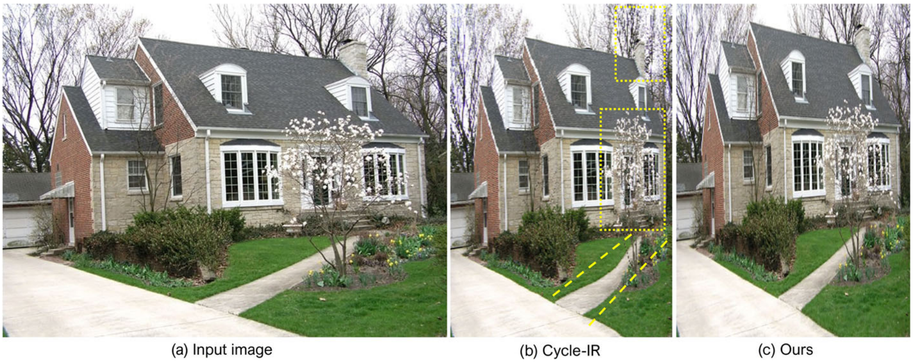


Fig. 16 Comparison with Cycle-IR

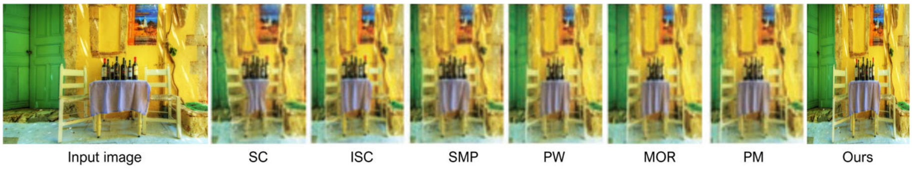


Fig. 17 Comparisons with content-aware and semantic-aware retargeting approaches

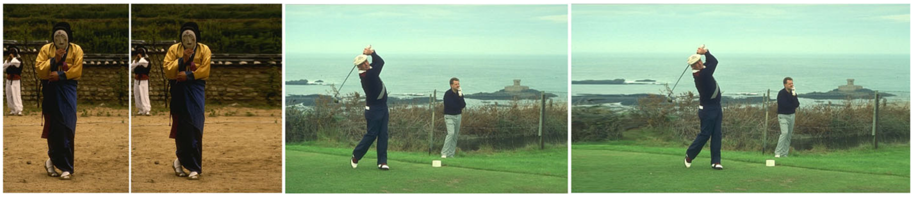


Fig. 18 Enlarging results. In each pair, left: input image, right: enlarged one

Appendix B: List of notations

Symbol	Definition
$\mathcal{I}$	Input image
SC	Seam carving operator
OMap	The energy map generated by TFS-Net
BMap	The energy map generated by (4)
A2R-Map	The final importance map generated by our model
SOD	Salient Object Detection
TFS-Net	The network we proposed to generate OMap
TFS	Feature Sharing Session module
AFS	Adjacent-layer Feature Sharing module
$\mathcal{X}^i$	Feature maps at layer i^{th}
$\mathcal{X}_u$	Feature maps at upper layer
$\mathcal{X}_l$	Feature maps at lower layer
$\mathcal{A}^s$	The source image/video in general
$\mathcal{A}^t$	The target image/video $\mathcal{A}^s$ after retargeting process
$\mathcal{P}$	A certain resizing operator
$\mathcal{R}$	A retargeting system using operator $\mathcal{P}$ to resize $\mathcal{A}^s$ and output $\mathcal{A}^t$
$\mathcal{M}$	An off-the-shelf method that $\mathcal{R}$ uses to define the importance in the input $\mathcal{A}^s$

Acknowledgements This work was supported in part by the National Science and Technology Council (under nos. 111-2221-E-006-112-MY3, 110-2221-E-006-135-MY3, 112-2221-E-019-063-MY3 and 110-2221-E-019-052-MY3), Republic of China (ROC), Taiwan. And this work is also supported by National Natural Science Foundation of China under Nos. U20B2070 and 61832016.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest. A part of the datasets generated and/or analysed during the current study are available in https://people.csail.mit.edu/mrub/retargetme, https://www.ee.nthu.edu.tw/cwlin/Retargeting_Quality/NRID.html, and https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/.

References

1. Suh B, Ling H, Bederson BB, Jacobs DW (2003) Automatic thumbnail cropping and its effectiveness. In: Proceedings of the 16th annual ACM symposium on user interface software and technology, pp 95–104
2. Kopf S, Guthier B, Lemelson H, Effelsberg W (2009) Adaptation of web pages and images for mobile applications. In: Multimedia on Mobile Devices 2009, vol. 7256, p 72560. International Society for Optics and Photonics
3. Avidan S, Shamir A (2007) Seam carving for content-aware image resizing. In: ACM SIGGRAPH 2007 Papers, p 10
4. Rubinstein M, Shamir A, Avidan S (2009) Multi-operator media retargeting. ACM Trans Graph (TOG) 28(3):1–11
5. Pritch Y, Kav-Venaki E, Peleg S (2009) Shift-map image editing. In: 2009 IEEE 12th international conference on computer vision, pp 151–158. IEEE
6. Lin S-S, Yeh I-C, Lin C-H, Lee T-Y (2012) Patch-based image warping for content-aware retargeting. IEEE Trans Multimed 15(2):359–368
7. Lin S-S, Lin C-H, Yeh I-C, Chang S-H, Yeh C-K, Lee T-Y (2013) Content-aware video retargeting using object-preserving warping. IEEE Trans Vis Comput Graph 19(10):1677–1686
8. Asheghi B, Salehpour P, Khiavi AM, Hashemzadeh M (2022) A comprehensive review on content-aware image retargeting: From classical to state-of-the-art methods. Signal Processing 108496
9. Cho D, Park J, Oh T-H, Tai Y-W, So Kweon I (2017) Weakly-and self-supervised learning for content-aware deep image retargeting. In: Proceedings of the IEEE international conference on computer vision, pp 4558–4567
10. Tan W, Yan B, Lin C, Niu X (2019) Cycle-ir: Deep cyclic image retargeting. IEEE Trans Multimed
11. Kajiura N, Kosugi S, Wang X, Yamasaki T (2020) Self-play reinforcement learning for fast image retargeting. In: Proceedings of the 28th ACM international conference on multimedia, pp 1755–1763
12. Lin J, Zhou T, Chen Z (2019) Deepir: A deep semantics driven framework for image retargeting. In: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp 54–59. IEEE
13. Kiess J, Kopf S, Guthier B, Effelsberg W (2018) A survey on content-aware image and video retargeting. ACM Trans Multimed Comput Commun Appl (TOMM) 14(3):1–28
14. Li X, Ling H (2009) Learning based thumbnail cropping. In: 2009 IEEE International Conference on Multimedia and Expo, pp 558–561. IEEE
15. Santella A, Agrawala M, DeCarlo D, Salesin D, Cohen M (2006) Gaze-based interaction for semi-automatic photo cropping. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 771–780
16. Guo D, Ding J, Tang J, Xu M, Zhao C (2015) Nif-based seam carving for image resizing. Multimedia Systems 21(6):603–613
17. Shen J, Wang D, Li X (2013) Depth-aware image seam carving. IEEE Trans Cybern 43(5):1453–1461
18. Wu J, Zhou W, Luo T, Yu L, Lei J (2021) Multiscale multilevel context and multimodal fusion for rgb-d salient object detection. Signal Processing 178:107766
19. Choi J, Kim C (2016) Sparse seam-carving for structure preserving image retargeting. J Signal Process Syst 85(2):275–283

20. Battiato S, Farinella GM, Puglisi G, Ravi D (2014) Saliency-based selection of gradient vector flow paths for content aware image resizing. IEEE Trans Image Process 23(5):2081–2095
21. Cui J, Cai Q, Lu H, Jia Z, Tang M (2020) Distortion-aware image retargeting based on continuous seam carving model. Signal processing 166:107242
22. Zhang X, Hu Y, Rajan D (2013) Dynamic distortion maps for image retargeting. J Vis Commun Image Represent 24(1):81–92
23. Guo Y, Liu F, Shi J, Zhou Z-H, Gleicher M (2009) Image retargeting using mesh parametrization. IEEE Trans Multimed 11(5):856–867
24. Wang Y-S, Tai C-L, Sorkine O, Lee T-Y (2008) Optimized scale-and-stretch for image resizing. In: ACM SIGGRAPH Asia 2008 Papers, pp 1–8
25. Zhang G-X, Cheng M-M, Hu S-M, Martin RR (2009) A shape-preserving approach to image resizing. In: Computer Graphics Forum, vol 28, pp 1897–1906. Wiley Online Library
26. Jin Y, Liu L, Wu Q (2010) Nonhomogeneous scaling optimization for realtime image resizing. Vis Comput 26(6):769–778
27. Niu Y, Liu F, Li X, Gleicher M (2012) Image resizing via non-homogeneous warping. Multimed Tools Appl 56(3):485–508
28. Hu W, Luo Z, Fan X (2014) Image retargeting via adaptive scaling with geometry preservation. IEEE J Emerg Sel Top Circ Syst 4(1):70–81
29. Panozzo D, Weber O, Sorkine O (2012) Robust image retargeting via axis-aligned deformation. In: Computer Graphics Forum, vol 31, pp 229–236. Wiley Online Library
30. Tan W, Yan B, Li K, Tian Q (2015) Image retargeting for preserving robust local feature: Application to mobile visual search. IEEE Trans Multimed 18(1):128–137
31. Kim Y, Jung S, Jung C, Kim C (2018) A structure-aware axis-aligned grid deformation approach for robust image retargeting. Multimed Tools Appl 77(6):7717–7739
32. Kim Y, Eun H, Jung C, Kim C (2018) A quad edge-based grid encoding model for content-aware image retargeting. IEEE Trans Vis Comput Graph 25(12):3202–3215
33. Liu S, Wei Z, Sun Y, Ou X, Lin J, Liu B, Yang M-H (2018) Composing semantic collage for image retargeting. IEEE Trans Image Process 27(10):5032–5043
34. Guo G, Wang H, Shen C, Yan Y, Liao H-YM (2018) Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression. IEEE Trans Multimed 20(8):2073–2085
35. Song E, Lee M, Lee S (2018) Carvingnet: content-guided seam carving using deep convolution neural network. IEEE Access 7:284–292
36. Wang Z, Zhang W, Zhou H (2019) Perception-guided multi-channel visual feature fusion for image retargeting. Signal Process Image Commun 79:63–70
37. Ahmadi M, Karimi N, Samavi S (2021) Context-aware saliency detection for image retargeting using convolutional neural networks. Multimed Tools Appl 80(8):11917–11941
38. Zhou Y, Chen Z, Li W (2020) Weakly supervised reinforced multi-operator image retargeting. IEEE Trans Circ Syst Video Technol 31(1):126–139
39. Shafieyan F, Karimi N, Mirmahboub B, Samavi S, Shirani S (2017) Image retargeting using depth assisted saliency map. Signal Process Image Commun 50:34–43
40. Li B, Duan L-Y, Lin C-W, Huang T, Gao W (2015) Depth-preserving warping for stereo image retargeting. IEEE Trans Image Process 24(9):2811–2826
41. Zhang W, Yao T, Zhu S, Saddik AE (2019) Deep learning-based multimedia analytics: a review. ACM Trans Multimed Comput Commun Appl (TOMM) 15(1s):1–26
42. Zhang Z, Lin H, Zhao X, Ji R, Gao Y (2018) Inductive multi-hypergraph learning and its application on view-based 3d object classification. IEEE Trans Image Process 27(12):5957–5968
43. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Show, attend and tell: Neural image caption generation with visual attention. In: International Conference on Machine Learning, pp 2048–2057. PMLR
44. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
45. Kanopoulos N, Vasanthavada N, Baker RL (1988) Design of an image edge detection filter using the sobel operator. IEEE Journal of solid-state circuits 23(2):358–367
46. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H-Y (2010) Learning to detect a salient object. IEEE Trans Pattern Anal Mach Intell 33(2):353–367
47. Goferman S, Zelnik-Manor L, Tal A (2011) Context-aware saliency detection. IEEE Trans Pattern Anal Mach Intell 34(10):1915–1926

48. Grundmann M, Kwatra V, Han M, Essa I (2010) Efficient hierarchical graph-based video segmentation. In: 2010 Ieee Computer society conference on computer vision and pattern recognition, pp 2141–2148. IEEE
49. Patel D, Nagar R, Raman S (2019) Reflection symmetry aware image retargeting. Pattern Recogn Lett 125:179–186
50. Cheng M-M, Mitra NJ, Huang X, Torr PH, Hu S-M (2014) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
51. Qin X, Zhang Z, Huang C, Gao C, Dehghan M, Jagersand M (2019) Basnet: Boundary-aware salient object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7479–7489
52. Qin X, Dai H, Hu X, Fan D-P, Shao L, Van Gool L (2022) Highly accurate dichotomous image segmentation. In: European Conference on Computer Vision, pp 38–56. Springer
53. Liu J-J, Hou Q, Cheng M-M (2020) Dynamic feature integration for simultaneous detection of salient object, edge, and skeleton. IEEE Trans Image Process 29:8652–8667
54. Tang F, Dong W, Meng Y, Ma C, Wu F, Li X, Lee T-Y (2019) Image retargetability. IEEE Trans Multimed 22(3):641–654
55. Zhang Y, Lin W, Zhang X, Fang Y, Li L (2016) Aspect ratio similarity (ars) for image retargeting quality assessment. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 1080–1084. IEEE
56. Liu C, Yuen J, Torralba A (2010) Sift flow: Dense correspondence across scenes and its applications. IEEE Trans Pattern Anal Mach Intell 33(5):978–994
57. Rubinstein M, Gutierrez D, Sorkine O, Shamir A (2010) A comparative study of image retargeting. ACM Trans Graph (Proc. SIGGRAPH ASIA) 29(6):160–116010
58. Simakov D, Caspi Y, Shechtman E, Irani M (2008) Summarizing visual data using bidirectional similarity. In: 2008 IEEE conference on computer vision and pattern recognition, pp 1–8. IEEE
59. Ma L, Lin W, Deng C, Ngan KN (2012) Image retargeting quality assessment: A study of subjective scores and objective metrics. IEEE J Sel Top Signal Process 6(6):626–639
60. Zhang L, Li X, Nie L, Yan Y, Zimmermann R (2016) Semantic photo retargeting under noisy image labels. ACM Trans Multimed Comput Commun Appl (TOMM) 12(3):1–22
61. Rubinstein M, Shamir A, Avidan S (2008) Improved seam carving for video retargeting. ACM Trans Graph (TOG) 27(3):1–9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.