Two-directional Image Retargeting with Region-Aware Texture Enhancement

Dubok Park
dubok.park@samsung.com

Samsung Research,
Samsung Electronics,
Seoul, South Korea

Abstract

In this paper, a novel framework for image retargeting based on two-directional non-uniform scaling with region-aware texture enhancement is proposed. First, the importance map is estimated via saliency and depth information from an input image. Then, two-directional scaling functions are derived from importance map by 1D projection onto horizontal and vertical axis according to estimated scaling portion which controls the resizing rate of each direction. Finally, the target image is generated by two-directional non-uniform scaling with region-aware guidance map. Experimental results validate the proposed framework can achieve content-preserving and visually pleasing results compared to the conventional methods.

1 Introduction

Recently, multimedia can be displayed depending on various environments for advertising, entertainment, and customer's needs. Therefore, image retargeting has been regarded as an important task in the computer vision area. Image retargeting is an algorithm that resizes image contents with an aspect ratio adequately regardless of the shape or size of a screen as shown in Figure 1.

Initially, various image retargeting methods have been proposed based on content-aware cropping and seam carving. Content-aware cropping based methods [1, 2, 3, 4] adjust the aspect ratio of an image by including important regions. Although these are simple and powerful approaches to preserve the contents of an original image, important regions may be eliminated from the original image in case there are important regions in the boundary of an image. Approaches based on seam carving [5, 6, 7, 8, 9, 10, 11] insert or carve out an image operator called seam chosen by energy function. Avidan and Shamir [5] derive the seam based on image energy function which accounts for the significance of each pixel. However, their solution occasionally provides the visual artifacts when misestimated seams are existing on the object or significant regions. For this reason, Frankovich and Wong [8] propose an enhanced seam carving using the absolute energy cost functional that incorporating energy gradient information into the optimization framework.

Continuous non-uniform scaling based approaches [12, 13, 14, 52, 53] find an optimal geometric transform by a warping function or scaling function via an importance map among regions in an image. These approaches efficiently detect salient objects or patterns and continuously transform an input image into a target size unlike seam carving methods. Wang et al. [13] propose the scale-and-stretch warping method which iteratively computes optimal scaling factor for local regions and updates a warped image collaborated by edge and

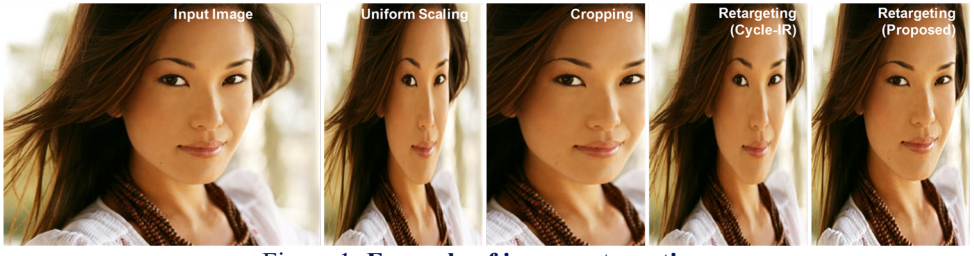


Figure 1: **Example of image retargeting.**

saliency map. However, their method sometimes fails to preserve the shapes of prominent image lines of arbitrary orientations.

Multi-operator based approaches [15, 16, 17, 18] exploit the combinations of various operators. Rubinstein et al. [15] combine seam carving with cropping and scaling in their architecture. However, multi-operator based methods dependently have a limitation of used operator. That is to say, those can have the problem of seam caving based methods if the seam carving is used as the operator.

Recently, deep-learning based methods [19, 20] have been extensively utilized to solve the aforementioned problems in image retargeting task. As with the advancements of technologies, video retargeting approaches [21, 22, 23, 24] also have been proposed for practical applications. Cho et al. [19] propose a weakly supervised network using a small object recognition dataset with twenty classes only. Thus, it has a limitation on representing the semantic contents of natural images. Tan et al. [20] propose a deep cyclic image retargeting approach, called Cycle-IR, to construct a single deep model without relying on any explicit user annotations. Although their approach provides compelling results in many cases, some failure results may occur if the important areas of source image are too large or scattered.

In this paper, a novel image retargeting framework is proposed with region-aware texture enhancement. The proposed method consists of three stages. First, the importance map is estimated from visual saliency, depth information, and input image. Two-directional scaling functions, then, are derived from the importance map with scaling ratio which is estimated by measure of statistical dispersion. Finally, input image is transformed by two-directional non-uniform scaling collaborating with region-aware guidance map to not only compensate degraded fine details, but also enhance overall quality.

The main contribution of this paper is first to transform an input image non-uniformly into the horizontal and vertical directions, two-directional way, according to the statistical distribution of scaling functions. Since the two-directional non-uniform scaling allows more input marginal regions which can be resized to be fitted for target aspect ratio, salient regions of input image can be preserved more efficiently and effectively compared to the unidirectional (horizontal or vertical only) non-uniform scaling approaches. Furthermore, retargeted images inherit a quality degradation at the non-significant regions since those regions are resized more than the important regions. To compensate quality degradation and enhance the fine details selectively, proposed approach exploits region-aware guidance map derived from second derivative kernels and inversed importance map. Figure 2 shows overall flow of the proposed method.

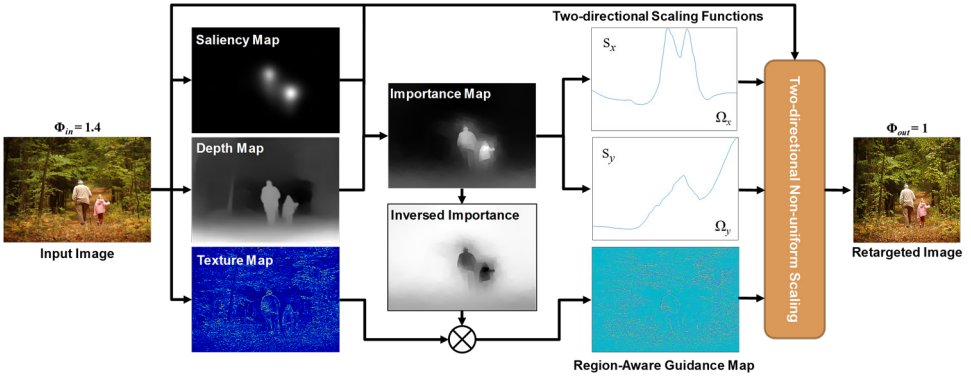


Figure 2: **Overview of the proposed approach.**

Figure 3: **Difference of proposed approach compared with the unidirectional non-uniform scaling.** (a) Conceptual example of unidirectional and two-directional non-uniform scaling, (b) Input image (1024x622, $\Phi_{in} \fallingdotseq 1.64$), (c) Cycle-IR [20] (512x622, $\Phi_{out} \fallingdotseq 0.82$), (d) Proposed approach (720x875, $\Phi_{out} \fallingdotseq 0.82$).

2 Related Work

Unidirectional Non-uniform Scaling for Retargeting. There are various approaches for solving an image retargeting problem such as content-aware cropping based approaches [1, 2, 3, 4], seam-carving based approaches [5, 6, 7, 8, 9, 10, 11], warping based approaches [12, 13, 14, 52, 53], hybrid multi-operator based approaches [15, 16, 17, 18], and deep-learning based methods [19, 20, 21, 23, 24]. However, most approaches utilize unidirectional non-uniform scaling for fitting target aspect ratio. Figure 3(a) shows the example case for decreasing aspect ratio of input image. For fitting target aspect ratio, conventional approach [20] reduces the width by half as shown in example case of Figure 3(c). In contrast, proposed approach reduces the width and enlarges the height for adjusting target aspect ratio. That is, proposed method can resize an input image non-uniformly with more freedom than existing approaches while preserving important regions as shown in Figure 3(d).

Importance Map. Early image retargeting approaches adopt an importance map derived from the saliency detection that models the human eye fixation process [25, 26, 27, 28, 29]. As these bottom-up methods are driven by low-level visual cues, edges and corners in images are detected rather than semantic regions. Recently, deep-learning based approaches [19, 20, 24, 30] have been actively developed to extract high-level semantic features of an input image. There are also approaches collaborating with depth information [31, 32, 33, 34, 35, 36] since the distance between the object and the camera is often associated with the object's significance. However, most approaches acquire the depth map using additional sensors or information such as Kinect or RGB-D sensors [31, 34, 35], stereo matching algorithms for stereo images [32, 33], and a user-provided relative depth map. In this regard,

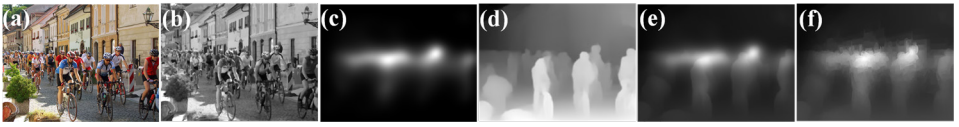


Figure 4: **Importance map estimation.** (a) Input image, (b) Median filtered grayscale input image, (c) Extracted saliency map, (d) Extracted depth map, (e) Combined depth and saliency map ($\lambda = 0.75$), (f) Estimated importance map optimized via WLS solution.

proposed method estimates importance map combining with saliency and depth information using a single input image for applying practical applications.

3 The Proposed Method

In Section 3.1 presents, first, a method for extracting the importance map. Then, Section 3.2 describes the procedure of deducing two-directional scaling functions. In Section 3.3, two-directional non-uniform scaling with region-aware guidance map is described.

3.1 Importance Map

In order to measure importance, proposed method takes into account not only visual saliency but also distance in the scene. To this end, proposed method adopts UNISAL [37] which achieves state-of-the-art performance on all video saliency datasets and is on par with the state-of-the-art for image saliency datasets, despite faster runtime and a 5 to 20-fold smaller model size compared to all competing deep methods. MiDas [38] is also chosen for generic monocular depth estimation since it has evaluated the robustness and generality of models via zero-shot cross-dataset transfer and improves performance across diverse environments. Figure 4(c) and (d) show the estimated saliency and depth map via aforementioned methods. The initial importance map A_{init} is defined as:

$$A_{init}(i,j) = \lambda \cdot B(i,j) + (1-\lambda) \cdot Q(i,j), \tag{1}$$

where, $B(i,j)$, $Q(i,j)$ are the saliency and depth map at location (i,j), respectively, and λ is a parameter for controlling balances between the saliency and the depth term. The initial importance map provides the significant regions of input image according to the human eye fixation distribution with distance priority as shown in Figure 4(e). However, A_{init} needs to be refined for improving the accuracy of importance map since it is derived by simple pixel addition without consideration of object details or shape of input image. Thus, further refining process is desirable. There have been several refinement filters suggested so far. The representatives are bilateral filter [47], guided filter [48], and WLS (Weighted Least Squares) filter [39], to name a few. WLS filter is chosen here for its halo removal effect while retaining shape of input image despite of its large computational complexity (more details are presented in supplementary). A_{init} is optimized using WLS solution cooperating with median filtered input image which contains large gradient and shape of input image while alleviating local noise. Using matrix notation, objective function for WLS solution minimizes the following equation:

$$(\mathbf{A}-\mathbf{A}_{init})^T(\mathbf{A}-\mathbf{A}_{init}) + \alpha(\mathbf{A}^T\mathbf{D}_x^T\mathbf{P}_x\mathbf{D}_x\mathbf{A} + \mathbf{A}^T\mathbf{D}_y^T\mathbf{P}_y\mathbf{D}_y\mathbf{A}), \tag{2}$$

where **A** is the optimized importance map, α is a constant value to control the smoothing rate, $\mathbf{D}_x$ and $\mathbf{D}_y$ are discrete differentiation operators, respectively, and the matrices $\mathbf{P}_x$, $\mathbf{P}_y$ are diagonal matrices containing the smoothness weights $P_x(i,j)$, $P_y(i,j)$ which defined by:

$$P_x(i,j)=\left(\left|\frac{\partial I_{med}(i,j)}{\partial x}\right|^{\beta}+\varepsilon\right)^{-1},\ P_y(i,j)=\left(\left|\frac{\partial I_{med}(i,j)}{\partial y}\right|^{\beta}+\varepsilon\right)^{-1}, \tag{3}$$

where, I_{med} is an image processed by median filtering from grayscale input image. Although these weights serve the purpose of smoothing the initial importance map to minimize the effect of local textures, these weights are decreased to small values when there is a large image contrast of gradient due to differences in distances. Exponent β determines the sensitivity to the gradients of I_{med} and ε is a small constant to prevent division by zero in (3).

Final optimized importance map **A** can be expressed via WLS solution, given by:

$$\mathbf{A}=\left(\mathbf{U}+\alpha(\mathbf{D}_x^T\mathbf{P}_x\mathbf{D}_x+\mathbf{D}_y^T\mathbf{P}_y\mathbf{D}_y\right)^{-1}\mathbf{A}_{init}, \tag{4}$$

where, **U** is the identity matrix of size $\mathbf{A}_{init}$. Due to the smoothness weights, the initial importance map can be optimized similarly to the input image while preserving energy of $\mathbf{A}_{init}$ as shown in Figure 4(f).

(a) (b) (c) (d) (e)

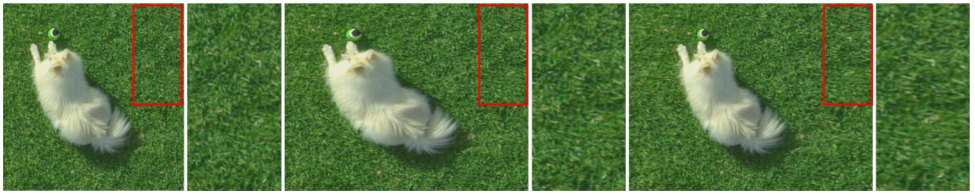
Figure 5: **Bidirectional scaling functions and ratio.** (a) Input image (400×300), (b) Importance map, (c) Retargeted image (339×339), (d) Horizontal scaling function s_x (φ_x = 0.6139, Ω_x = 339), (e) Vertical scaling function s_y (φ_y = 0.3861, Ω_y = 339).

3.2 Two-directional Scaling Functions

Using importance map, input image can be transformed the target aspect ratio. The aspect ratio of image Φ can be expressed as:

$$\Phi=\frac{w}{h}, \tag{5}$$

where, w and h are image width and height, respectively. While unidirectional scaling controls the width or the height only, proposed two-directional scaling adjusts the both direction (width, height) as shown in Figure 3(a). To determine ratio in each direction, proposed method exploits the distribution of scaling function [24] which represents the scaling factor at certain location of the input image as shown in Figure 5(d) and (e). Two-directional scaling functions can be expressed as follows:



(a) Input image (b) Uniform scaling (c) Non-uniform scaling

Figure 6: **Quality degradation in non-uniform scaling.**

$$s_x(i) = \frac{\sum_j A(i,j)}{\sum_i \sum_j A(i,j)} \Omega_x, \quad s_y(j) = \frac{\sum_i A(i,j)}{\sum_j \sum_i A(i,j)} \Omega_y, \tag{6}$$

where, s_x and s_y are horizontal and vertical scaling function, respectively, Ω_x and Ω_y denote the width and height of target image, which are expressed as follows:

$$\Omega_x = w \pm \varphi_x \Delta, \quad \Omega_y = h \mp \varphi_y \Delta, \tag{7}$$

where, φ_x and φ_y are scaling ratio of horizontal and vertical, respectively, Δ is entire scaling factor. Note that, $\varphi_x\Delta$ and $\varphi_y\Delta$ have the opposite sign as shown in Figure 3(a). To determine optimal ratios of φ_x and φ_y, proposed approach exploits distribution of scaling functions since it denotes the region-preserving factor in each direction. In the example of Figure 5, it is reasonable to resize the horizontal direction more than vertical direction since the significant region of s_x is more concentrated than s_y. To this end, using entropy-based dispersion measure [40] which quantifies how "evenly" is the probability distributed, proposed method defines two-directional scaling ratios in each direction as follows:

$$\varphi_x = \frac{\rho_x}{\rho_x + \rho_y}, \quad \varphi_y = \frac{\rho_y}{\rho_x + \rho_y}, \tag{8}$$

where,

$$\rho_x = \exp\left[-C_h(s_x(i))\right], \quad \rho_y = \exp\left[-C_h(s_y(j))\right], \tag{9}$$

where $C_h(\cdot)$ is the relative entropy-based dispersion coefficient [40]. In the example of Figure 5(d) and (e), φ_y is smaller than φ_x since s_y is more distributed overall than s_x. Given the target aspect ratio Φ_{out}, entire scaling factor Δ can be expressed as:

$$\Delta = \begin{cases} \dfrac{h \cdot \Phi_{out} - w}{\varphi_x + \varphi_y \cdot \Phi_{out}} & , \Phi_{out} > \Phi_{in} \\ -\dfrac{h \cdot \Phi_{out} - w}{\varphi_x + \varphi_y \cdot \Phi_{out}} & , \Phi_{out} < \Phi_{in} \end{cases} \tag{10}$$

where, Φ_{in} and Φ_{out} are the input and target aspect ratio, respectively. In the example case of Figure 5, Δ is equal to 100, and if $\varphi_x\Delta$ and $\varphi_y\Delta$ are floating point numbers, they will round off to integer number. In order to provide a better insight into the two-directional scaling function, ablation experiments are presented in supplementary to show how scaling ratio affects retargeted result.

Figure 7: **Illustration of a region-aware guidance map.** (a) Input image, (b) Retargeted image without region-aware guidance, (c) Retargeted image with region-aware guidance (d) Texture map, (e) Region-aware guidance map.

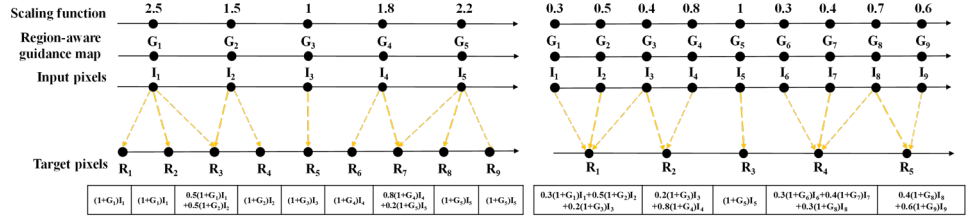


Figure 8: **Example of region-aware target pixel generation via non-uniform scaling.** Upscaling case (left) and downscaling case (right).

3.3 Region-aware Guidance Map & Two-directional Non-uniform Scaling

With two-directional scaling functions, target image can be generated from input image via non-uniform scaling. However, quality of target image may be degraded since the non-uniform scaling adjusts the unimportance region (grass) more than importance area (dog) as shown in Figure 6 (more results are presented in supplementary). To improve the quality of target image, proposed method resizes input image collaborating with region-aware guidance map which is defined using texture map [41] and inversed importance map as follows:

$$G(i,j) = (1 - A(i,j)) \cdot T(i,j), \tag{11}$$

where $G(i,j)$ is the region-aware guidance map, $T(i,j)$ is the texture map extracted by second derivative kernels [41]. As shown in Figure 7, significant region (dog) of the region-aware guidance map is close to zero which means this region maintains original texture. On the contrary, unimportant region (grass) in $G(i,j)$ has higher or lower value which means this region enhances the fine details according to the rate of gradient change. Note that, $G(i,j)$ is ranged between -1 and 1 due to the algebraic function [41] as shown in Figure 7(e).

Given the two-directional scaling functions and region-aware guidance map as shown in Figure 8, target image is generated from an input image via weighted sum of estimated parameters while conventional methods only exploit scaling function [24, 49]. Each target pixel is filled up by input pixels until stacked value of scaling function is equal to 1. Excessive or deficient value is used for the next pixel generation. In accordance with region-aware guidance map, target pixel is enhanced by a proportion to the value of region-aware guidance map. Algorithm 1 and 2 practically explain proposed framework in pseudocode with respect to the horizontal direction (vertical direction is the same manner).

Algorithm 1 Region-aware target pixel generation ($\Phi_{out} > \Phi_{in}$)

Input: $I(i, j)$, $G(i, j)$, and $s_x(i)$
Output: $T(i', j)$

$I(i, j)$: Input source image
$G(i, j)$: Region-aware guidance map
$s_x(i)$: Scaling function
$T(i', j)$: Target image ($i' > i$)
Ω_x: Target image width ($\Omega_x > w$)

function TARGETIMAGE_UP($I(i, j)$, $G(i, j)$, $s_x(i)$, Ω_x)
  $i = 1$
  for $i' = 1$: Ω_x **do**
    if $s_x(i) > 1$ **then**
      $T(i', j) = (1+G(i, j))\cdot I(i, j)$
      $s_x(i) \leftarrow s_x(i) - 1$
    else
      $T(i', j) = s_x(i)\cdot(1+G(i, j))\cdot I(i, j) + (1 - s_x(i))\cdot(1+G(i+1, j))\cdot I(i+1, j)$
      $s_x(i+1) \leftarrow s_x(i)+s_x(i+1) - 1$
      $i \leftarrow i+1$
    end
  end for
end function

Algorithm 2 Region-aware target pixel generation ($\Phi_{out} < \Phi_{in}$)

Input: $I(i, j)$, $G(i, j)$, and $s_x(i)$
Output: $T(i', j)$

$T(i', j)$: Target image ($i' < i$)
Ω_x: Target image width ($\Omega_x < w$)

function TARGETIMAGE_DOWN($I(i, j)$, $G(i, j)$, $s_x(i)$, Ω_x)
  $i = 1$
  for $i' = 1$: Ω_x **do**
    temp_s_x = $s_x(i)$
    while temp_s_x < 1 **then**
      if temp_s_x + $s_x(i+1) > 1$ **then**
        $T(i', j) \leftarrow T(i', j) + (1 - \text{temp\_}s_x)\cdot(1+G(i+1, j))\cdot I(i+1, j)$
        $s_x(i+1) \leftarrow$ temp_s_x + $s_x(i+1) - 1$
        break
      else
        $T(i', j) \leftarrow T(i', j) + s_x(i)\cdot(1+G(i, j))\cdot I(i, j)$
        temp_s_x ← temp_s_x + $s_x(i+1)$
        $i \leftarrow i+1$
      end
    end while
  end for
end function

Figure 9: **Representative results of proposed method with arbitrary sizes.**

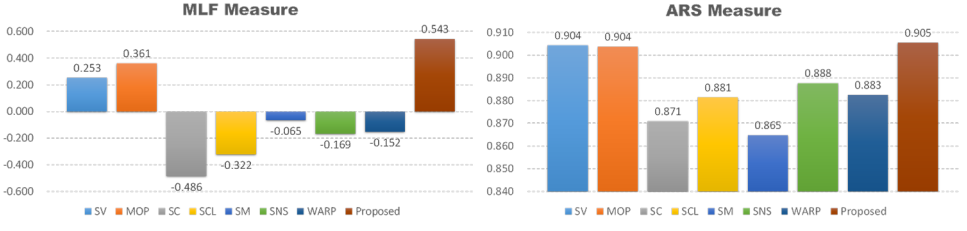


Figure 10: **Performance comparison of MLF [45] and ARS [50] on RetargetME [46]**

4 Experiments and Discussion

In this section, the performance of proposed method is compared and discussed with experimental results of conventional methods. To validate effectiveness qualitatively and quantitatively, proposed method is compared with existing retargeting methods such as Streaming Video (SV) [42], Multi-operator (MOP) [15], Seam Carving (SC) [6], Shift Maps (SM) [43], Scale-and-Stretch (SNS) [13], WARP [44], Weakly- and Self-Supervised (WSS) [19], and Cycle-IR [20]. RetargetME [46] dataset is used for experiments since it is a benchmark for image retargeting and can be download from project webpage easily. In the experiments, λ, α, and β were set to 0.75, 1, and 1.2, respectively. Figure 9 shows the

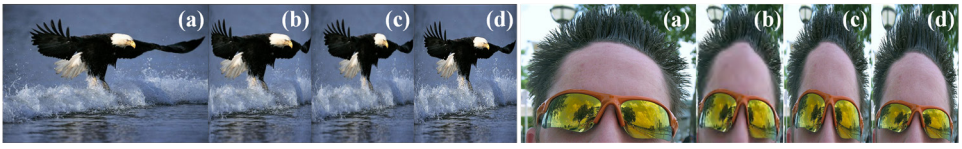


Figure 11: **Visual comparisons with deep-learning based approaches ($\Phi_{out} = 0.5\times\Phi_{in}$).** (a) Input image, (b) WSS [19], (c) Cycle-IR [20], (d) Proposed method.

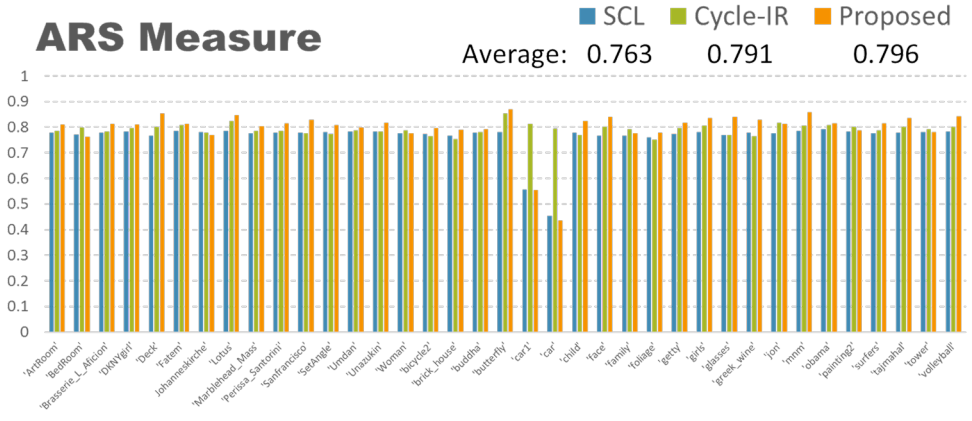


Figure 12: **Performance comparison of ARS measure [50] ($\Phi_{out} = 0.5\times\Phi_{in}$).**

representative results of proposed approach with respect to various target aspect ratios. In spite of large scale range (from 0.5 to 1.5), proposed approach provides well-preserved contents (such as butterfly, and people) and visually pleasing results. For the objective evaluation, MLF-based quality measure [45] and Aspect Ratio Similarity (ARS) [50] are used to evaluate performance of proposed method with respect to SV, MOP, SC, Simple Scaling (SCL), a.k.a. uniform scaling, SM, SNS, WARP. Note that, the MLF and ARS have better quality as it becomes larger. Figure 10 shows the comparison results of MLF and ARS on RetargetME dataset. Each image of the comparison results is retargeted with 25% or 50% of width change using the selected operators. Each score in MLF and ARS is average value for 37 datasets. In part of result, SC shows the discontinuities which are due to the pixels being excessively removed, which results in a lower score than the others. Although SV and MOP show higher scores than conventional methods, proposed method provides the highest average values in MLF and ARS measures. Visual comparison with conventional methods can be found in the supplementary.

Proposed method is also compared to deep-learning based techniques as shown in Figure 11. For "eagle" image, proposed method preserves better for hawk's wings and head than conventional methods. For "glass" image, WSS shows some distortion in men's forehead and side of head is reduced too much. Although Cycle-IR provides good result without distortion, proposed method preserves glass shape better than Cycle-IR. For the objective evaluation, ARS is used to validate effectiveness of the proposed method with respect to SCL, Cycle-IR (results of Cycle-IR, $\Phi_{out} = 0.5\times\Phi_{in}$, were uploaded on GitHub [51]). Figure 12 shows the comparison result of ARS measure with respect to RetargetME dataset. Although Cycle-IR shows better results on 'car1' and 'car', proposed method provides better performance overall as shown in average score of Figure 12. Qualitative comparison in part of result can be found in the supplementary.

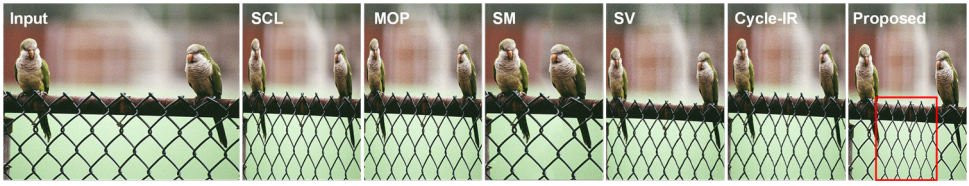


Figure 13: **Some failure case ($\Phi_{out} = 0.5 \times \Phi_{in}$).**

Although the proposed method is capable of producing compelling target images, proposed method has limitation cases as shown in Figure 13. This is due to the fact that proposed approach has no regard for global consistency. Thus, patch coherence between the input and target images may help the proposed approach to achieve better performance.

5 Conclusion

In this paper, a novel framework for image retargeting is proposed via two-directional non-uniform scaling with region-aware guidance map. The proposed method first estimates the importance map by collaborating saliency, depth map, and median filtered input image. Then, two-directional scaling functions are derived from importance map with measure of statistical dispersion. Finally, target image is generated by two-directional non-uniform scaling cooperating with region-aware guidance map. From the limited set of experiments, it was shown to successfully generate retargeted image without artifacts while retaining details. The future works of proposed method may include handling extremely aspect ratio change and consideration of temporal coherence for video retargeting.

References

[1] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen. GazeBased Interaction for SemiAutomatic Photo Cropping. In *Proc. of the SIGCHI conference on Human Factors in computing systems*, pp. 771-780, April 2006.

[2] F. Liu, and M. Gleicher. Video Retargeting: Automating Pan and Scan. In *Proc. of the 14th ACM international conference on Multimedia*, pp. 241-250, Oct. 2006.

[3] T. Deselaers, P. Dreuw, and H. Ney. Pan, Zoom, Scan - Time-coherent, Trained Automatic Video Cropping. In *CVPR*, 2008.

[4] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based Photo Cropping. In *Proc. of the 17th ACM international conference on Multimedia*, pp. 669-672, Oct. 2009.

[5] S. Avidan, and A. Shamir. Seam Carving for Content-Aware Image Resizing. *ACM Trans. on Graph.* (*ToG*), vol. 26, issue 3, July 2007.

[6] M. Rubinstein, A. Shamir, S. Avidan. Improved Seam Carving for Video Retargeting. *ACM transactions on graphics (ToG)*, vol. 27, issue 3, Aug. 2008.

[7] V. Bucha, I. Safonov, M. Rychagov, J.K. Hong, and S.H. Kim. Retargeting of digital images and documents. In *Proc. SPIE 7250,* Digital Photography V, 72500Z, January 2009.

[8] M. Frankovich, and A. Wong. Enhanced Seam Carving via Integration of Energy Gradient Functionals. *IEEE Signal Processing Letters*, vol. 18, issue 6, pp. 375-378, June 2011.

[9] P. Razzaghi, and S. Samavi. Image retargeting using nonparametric semantic segmentation. *Multimedia Tools and Applications*, vol. 74, issue 24, pp. 11517-11536, Dec. 2015.

[10] C. Li, R. Hu, C. Liang, C. Xiao and W. Ruan. Faster Seam Carving for Video Retargeting. In *ICIP*, pp. 823–827, Oct. 2018.

[11] R. Furuta, I. Tsubaki, and T. Yamasaki. Fast Volume Seam Carving with Multi-pass Dynamic Programming. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, issue 5, pp. 1087-1101, May 2018.

[12] S. S. Lin, C. H. Lin, I. C. Yeh, S. H. Chang, C. K. Yeh, and T. Y. Lee. Content-Aware Video Retargeting Using Object-Preserving Warping. *IEEE transactions on visualization and computer graphics*, vol. 19, issue 10, pp. 1677-1686, 2013

[13] Y.S. Wang, C. L. Tai, O. Sorkine, and T. Y. Lee. Optimized Scale-and-stretch for Image Resizing. *ACM Trans. on Graph.* (*ToG*), vol. 27, issue 5, 2008.

[14] C. P. Lau, C. P. Yung, and L. M. Lui. Image retargeting via Beltrami representation. *IEEE Transactions on Image Processing*, vol. 27, issue 12, pp. 5787-5801, July 2018.

[15] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. *ACM Trans. on Graph.* (*ToG*), vol. 28, no. 3, pp. 23 :1–23:11, 2009.

[16] W. Dong, N. Zhou, T. Y. Lee, F. Wu, Y. Kong, and X. Zhang. Summarization-Based Image Resizing by Intelligent Object Carving. *IEEE transactions on visualization and computer graphics*, vol. 20, issue 1, pp.111-124, January 2014.

[17] W. Y. Yo, J. J. Leou, and H. H. Hsiao. Video Retargeting Using Non-homogeneous Scaling and Cropping. In *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1-5, January 2014.

[18] W. Dong, F. Wu, Y. Kong, X. Mei, T. Y. Lee, and X. Zhang. Image Retargeting by Texture-Aware Synthesis. *IEEE Transactions on Visualization & Computer Graphics*, vol. 22, issue 2, pp. 1088-1101, Feb. 2016.

[19] D. Cho, J. Park, T. H. Oh, Y. W. Tai, and I. S. Kweon. Weakly- and Self-Supervised Learning for Content-Aware Deep Image Retargeting. In *Proc*. *ICCV*, pp. 4568-4577, Oct. 2017.

[20] W. Tan, B. Yan, C. Lin, and X. Niu. Cycle-IR: Deep Cyclic Image Retargeting. *IEEE Transactions on Multimedia*, vol. 22, issue 7, pp. 1730-1743, July 2020.

[21] S. I. Cho, and S. J. Kang. Extrapolation-Based Video Retargeting with Backward Warping Using an Image-to-Warping Vector Generation Network. *IEEE Signal Processing Letters*, vol. 27, pp. 446-450, Feb. 2020.

[22] R. Furuta, I. Tsubaki, and T. Yamasaki. Fast Volume Seam Carving With Multipass Dynamic Programming. *IEEE Transactions on Consumer Electronics*, vol. 57, issue 2, May 2011.

[23] S. I. Cho, and S. J. Kang. Temporal Incoherence-Free Video Retargeting Using Foreground Aware Extrapolation. *IEEE Transactions on Image Processing*, vol.29, pp 4848-4861, March 2020.

[24] H. Nam, D. Park, and K. Jeon. Jitter-Robust Video Retargeting With Kalman Filter And Attention Saliency Fusion Network. In *Proc. ICIP*, 2020.

[25] Y. Ding, J. Xiao, and J. Yu. Importance filtering for image retargeting. In *Proc. CVPR*, Jun. 2011.

[26] J. Luo. Subject content-based intelligent cropping of digital photos. In *Proc. ICME*, Jul. 2007.

[27] L. Marchesotti, C. Cifarelli, and G. Csurka. A framework for visual saliency detection with applications to image thumbnailing. In *Proc. ICCV*, Sep./Oct. 2009.

[28] R. Zhao, W. Ouyang, H. Li, and X. Wang. Saliency detection by multicontext deep learning. In *Proc. CVPR*, Jun. 2015.

[29] J. Chen, G. Bai, S. Liang, and Z. Li. Automatic image cropping: A computational complexity study. In *Proc. CVPR*, Jun. 2016.

[30] B. Xiong, S. D. Jain, and K. Grauman. Pixel Objectness: Learning to Segment Generic Objects Automatically in Images and Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, issue 11, pp. 2677-2692, Nov. 2019.

[31] J. Shen, D. Wang, and X. Li. Depth-Aware Image Seam Carving. *IEEE Transactions on Cybernetics*, vol. 43, issue 5, pp. 1453-1461, Oct. 2013.

[32] T. D. Basha, Y. Moses, and S. Avidan. Stereo Seam Carving a Geometrically Consistent Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, issue 10, pp. 2513-2525, Oct. 2013.

[33] F. Shafieyan, N. Karimi, B. Mirmahboub, S. Samavi, and S. Shirani. Image seam carving using depth assisted saliency map. In *Proc. ICIP*, Oct. 2014.

[34] W. Y. Lin, C. F. Tsai, and P. C. Wu. Image retargeting using RGB-D camera. *Multimed. Tools Appl.*, 74, pp. 3155-3170, 2015.

[35] F. Shao, W. Lin, W. Lin, G. Jiang, M. Yu, and R. Fu. An Energy-Constrained Video Retargeting Approach for Color-Plus-Depth 3D Video. *Journal of Display Technology*, vol. 12, issue 5, pp. 491-499, 2016.

[36] F. Shafieyan, N. Karimi, B. Mirmahboub, S. Samavi, and S. Shirani. Image retargeting using depth assisted saliency map. *Signal Processing: Image Communication*, vol. 50, pp. 34-43, 2017.

[37] R. Droste, J. Jiao, and J. A. Noble. Unified Image and Video Saliency Modeling. In *Proc. ECCV*, 2020.

[38] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2020.3019967.

[39] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edgepreserving decomposition for multi-scale tone and detail manipulation. *ACM Trans. on Graph.* (*ToG*), vol. 27, issue 3, no. 67, Aug. 2008.

[40] L. Kostal, P. Lansky, and O. Pokora. Measures of statistical dispersion based on Shannon and Fisher information concepts. *Information Sciences*, vol. 235, pp. 214-223, 2013.

[41] D. Park. Edge-Guided Image Downscaling With Adaptive Filtering. In *Proc. ICIP*, 2020.

[42] P. Krahenbuhl, M. Lang, A. Hornung, and M. H. Gross. A system for retargeting of streaming video. *ACM Trans. Graph.* (*ToG*), vol. 28, no. 5, 2009.

[43] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. In *Proc. ICCV*, 2009.

[44] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *Proc. ICCV*, Oct. 2007.

[45] Y. Zhang, W. Lin, Q. Li, W. Cheng, and X. Zhang. Multiple-Level Feature-Based Measure for Retargeted Image Quality. *IEEE Transactions on Image Processing*, vol. 27, issue 1, Jan. 2018.

[46] M. Rubinstein, D. Gutierrez, O. Sorkine, and Ariel Shamir. A comparative study of image retargeting. *ACM Trans. on Graph.* (*ToG*), vol. 29, no. 6, Dec. 2010.

[47] C. Tomasi, R. Manduchi. Bilateral filtering for gray and color images. In *Proc. ICCV*, Jan. 1998

[48] K. He, J. Sun, and X. Tang. Guided Image Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2012.213.

[49] J. Kim, J. Kim, and C. Kim. Image and video retargeting using adaptive scaling function. In *Proc. EUSIPCO*, Aug. 2009.

[50] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and Leida Li. Backward Registration-Based Aspect Ratio Similarity for Image Retargeting Quality Assessment. *IEEE Transactions on Image Processing*, vol. 25, issue 9, pp. 4286-4297, Jun 2016.

[51] https://github.com/mintanwei/Cycle-IR

[52] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In *Proc. CVPR*, Jun. 2008.

[53] C. Barnes, E. Shechtman, A. Finkelstein, and D. B Goldman. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing. *ACM transactions on graphics (ToG)*, vol. 28, issue 3, Aug. 2009.