Prune and Repaint: Content-Aware Image Retargeting for any Ratio

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Image retargeting is the task of adjusting the aspect ratio of images to suit different display devices or presentation environments. However, existing retargeting methods often struggle to balance the preservation of key semantics and image quality, resulting in either deformation or loss of important objects, or the introduction of local artifacts such as discontinuous pixels and inconsistent regenerated content. To address these issues, we propose a content-aware retargeting method called PruneRepaint. It incorporates semantic importance for each pixel to guide the identification of regions that need to be pruned or preserved in order to maintain key semantics. Additionally, we introduce an adaptive repainting module that selects image regions for repainting based on the distribution of pruned pixels and the proportion between foreground size and target aspect ratio, thus achieving local smoothness after pruning. By focusing on the content and structure of the foreground, our PruneRepaint approach adaptively avoids key content loss and deformation, while effectively mitigating artifacts with local repainting. We conduct experiments on the public RetargetMe benchmark and demonstrate through objective experimental results and subjective user studies that our method outperforms previous approaches in terms of preserving semantics and aesthetics, as well as better generalization across diverse aspect ratios. Code will be publicly available.

1 Introduction

With the popularity of multi-screen and multi-aspect environments, people's demands for the adaptability and aesthetics of images across different devices are increasing. Consequently, image retargeting [35, 20, 9], which aims to adjust the aspect ratio to fit various display devices or presentation environments while preserving the key content and maintaining the quality of the images, has distinctive applications yet is understudied.

The core challenge of image retargeting lies in simultaneously 1) preserving the main information and 2) avoiding artifacts such as deformation and distortion on key objects. Intuitive solutions for this task include scaling and cropping. As shown in Figure 1 (b), scaling entirely preserves all contents but results in severe deformation, decreasing aesthetic appeal and image quality, making it difficult to recognize the figures. On the contrary, crop-based methods [38, 29] introduce no artifacts but often results in the loss of key semantics (see Figure 1 (c)). To relieve these problems, following methods typically use pixel-shifting operators. Noticeable works include seam-carving [1, 28], which calculates energy using manual operators [6, 2] to identify seams for deletion. Some other works [5, 39, 14] further integrate these traditional operators to enhance the generalization to different scenarios. However, as illustrated in Figure 1 (d), without semantics guidance for crucial regions, this line of methods often leads to content loss or distortion on important objects, as well as inconsistent pixels in the foreground.

Figure 1: An example to show bad cases such as deformation, content loss, discontinuity in lines, inconsistent results and a good case.

Regarding the power of deep learning tools, some methods [16, 34, 4] integrate deep semantic features to guide the deletion or preservation of pixels in traditional pixel-shifting methods. However, these methods fail to differentiate the semantic significance within objects (e.g., face is more important than hair in a person), thus often leading to object distortion when meets oversized objects. Moreover, these methods only focus on foreground regions, typically leading to discontinuous backgrounds and decreased aesthetic appeal. Towards better aesthetics and region consistency, other works [3, 31, 22, 8] achieve image retargeting from a generative perspective with Generative Adversarial Networks (GANs) [10]. These methods often implicitly learn the semantic distribution of images to regenerate retargeted images. Due to the absence of explicit semantic prior and the weakness of GANs in capturing the global data distribution [23], these methods will generate all regions without selection, resulting in inconsistent generation of key objects (Figure 1 (e)).

To tackle the issues mentioned above, we present a content-aware retargeting model that can maintain the essential semantics, their appearance consistency, and aesthetics while being adaptable to any aspect ratio. To alleviate semantic loss, we introduce content-aware seam-carving (CSC), which incorporates hierarchical semantic information induced from semantic/spatial saliency to differentiate the energy to perform scene-level (i.e., background and foreground) and object-level (i.e., components in the foreground) pruning, thereby maximizing the preservation of key objects and their discriminative semantic elements. To mitigate artifacts introduced by pixel removal, we further propose an adaptive repainting (AR) method based on diffusion models, consisting of an Adaptive Repainting Region Determination (ARRD) module and an Image-guided Repainting (IR) module.

The two modules work together to adaptively repaint scenes with varying foreground sizes. The ARDD module is responsible for determining which regions of the image need to be repainted. It does this by identifying abrupt pixels that have a high density of removed surrounding pixels. Then, it considers the foreground size and desired ratio to determine inpainting or outpainting. This approach ensures that important objects are preserved in the image even if they exceed the expected size, resulting in a flexible method that can handle images with any aspect ratio. Subsequently, IR refines the repainting process by using the original image as a reference to restore and repaint these regions. Compared to previous global generation methods [3, 31, 22] without maintaining foreground consistency with the original image, our approach selectively regenerates the abrupt pixels, preserving the foreground consistency and local smoothness effectively.

Our contributions can be summarized as follows:

1) We introduce a content-aware image retargeting framework that is applicable to any aspect ratio. By incorporating content-aware seam-carving, our approach enables pixel pruning with hierarchical semantic differentiation

2) We propose an adaptive repainting method that utilizes image-conditioned stable diffusion models. This method dynamically determines whether to inpaint or outpaint based on different aspect ratios, leading to local smoothness and aesthetically pleasing outcomes.

3) Through extensive experiments involving various aspect ratios, our method demonstrates superior performance compared to other approaches in terms of both objective and subjective evaluations. It excels in preserving object completeness, coherence, and generalization.

2 Related Work

2.1 Image Retargeting

Existing image retargeting methods revolve around two main themes: preserving the main information and avoiding artifacts. Early image retargeting methods often fail to balance these two aspects. For instance, scaling attempts to maintain overall elements by uniformly removing pixels but struggles with significant changes in aspect ratios, resulting in severe deformation of key objects. Cropping-based methods [38, 29] chooses the best window of target size from the original image, which preserves the structure but leads to the loss of crucial information outside the window. Seam-carving [1] attempts to balance content completeness and quality by calculating energy maps to remove lower-energy seams. However, due to the lack of semantics, when the background is complex, this method usually result in the distortion in foreground.

The rise of deep learning [15] has introduced semantic information to image retargeting. DeepIR [16] adopts pretrained VGG [32] to explicitly extract semantic information and retargets the image from a coarse semantic space to fine pixel space. SmartScale [4] utilizes existing object detection model to assist seam-carving. However, these methods ignore the semantic differences within important regions, resulting in deformation within the oversized regions. In addition, neglecting the background can lead to discontinuities in background pixels, thereby affecting the aesthetic appeal of the image. For aesthetics and local smoothness, some methods adopt Generative Adversarial Networks (GANs) [10] to generate the retargeted results. InGAN [31] and SinGAN [30] divide the image into patches and learn the internal distribution of patches, destroying the overall semantics. To training a GAN without partitioning the image, MRGAN [22] adopts multi-operator to generate a paired dataset, which is constrained by the handcraft, MCGAN [7] introduces mask to highlight importance areas. However, due to the limitations of implicit semantic expression, these methods preserve the global semantics but destroy the details, resulting in inconsistent appearance with the original image.

In contrast, our method is content-aware for selective pruning and adaptive repainting. It is able to maintain the key semantics and appearance while ensuring local smoothness and aesthetics, and it has stronger generalization for different aspect ratios.

2.2 Diffusion Models for Image Generation

Nowadays, diffusion models [11, 33, 26] have become the mainstream models for generative tasks due to their powerful ability to model complex distributions. Stable Diffusion [26] is the first generative model based on latent diffusion models. The progressively denoising diffusion in latent space significantly enhances the efficiency, stability, realism, and controllability of image generation. Subsequently, various improvements [24] and variations [37, 36] of stable diffusion models have been proposed. For instance, SDXL [24] adopts a larger backbone and finetunes it using a complicated dataset with multiple aspect ratios to improve its versatility.

However, such text-to-image (T2I) models are hard to generate complex scenes and achieve more detailed control, as a significant amount of text control is labor-intensive and the T2I models struggle to accurately comprehend numerous and complex text prompts. To tackle this issue, other conditioning methods [37, 36] are proposed. The introduction of ControlNet [37] expands the applications of Stable Diffusion with different image-based conditional control, including depth images, mask images, canny images, etc. IP-Adapter [36], a newly proposed image-to-image (I2I) model, introduces image prompts to control condition with an decoupled cross-attention adapter branch, highly enhancing the controllability of the generative image.

In our task, we introduce image-guided local repainting into image retargeting, which enjoys the advantage of more precise semantic preservation and more controllable local generation compared to global regeneration.

3 Method

3.1 Overall Architecture

The overall architecture is illustrated in Figure 2. Specifically, a saliency detection model is adopted to obtain the semantic saliency, which will further be combined with the initial energy map to guide

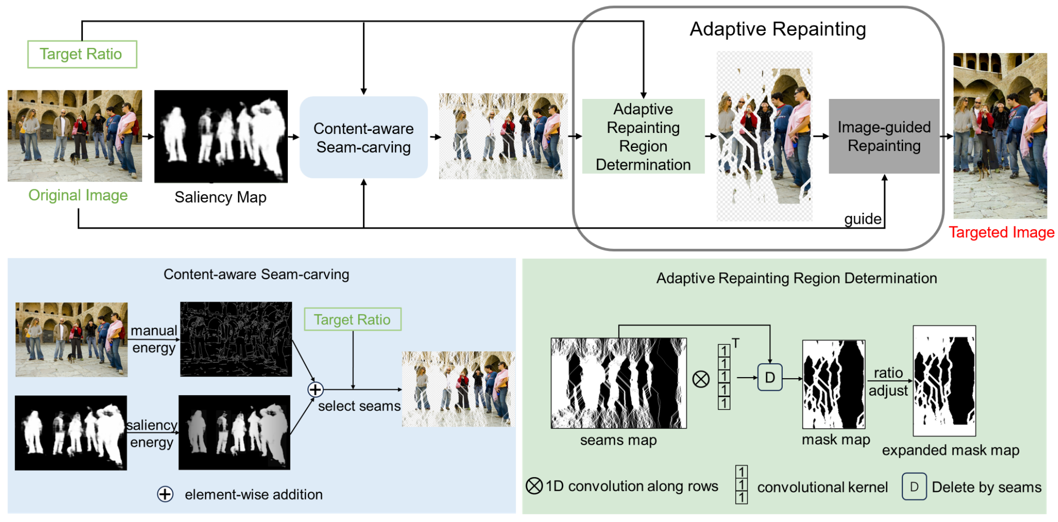


Figure 2: The overall architecture of our proposed PruneRepaint. The input consists of a RGB image and a target ratio. The saliency map, obtained by saliency detection, is further introduced into content-aware seam-carving module for preliminary retargeting. The preliminary retargeted result is then processed by the adaptive repainting region determination module to identify the abrupt pixel regions that need to be repainted. Utilizing the original image as guidance, the image is inpainted with the image-guided repainting module, generating the final targeted image with target ratio.

the determination of pruning pixels, making the pruning content-aware. After that, an adaptive repainting region determination module is applied to identify the abrupt pixels and determine the repainting regions, and an image-guided stable repainting module is further used to repaint them to output the final retargeted image.

3.2 Content-aware Seam-carving

Seam-carving is a typical pixel-shifting retargeting method, which calculates the energy of image and prioritizes deleting the seams with lower energy. For simplicity, we only discuss the scenario of deleting vertical seams in this section. The energy function in seam-carving is formulated as follows[1]:

$$Energy(I(x,y)) = |\frac{\partial}{\partial x}I(x,y)| + |\frac{\partial}{\partial y}I(x,y)|, \tag{1}$$

where $I(x,y)$ denotes the pixel at position (x,y) in the image. Seam-carving is often criticized for its lack of semantic information, which can lead to the distortion of key objects. To address this issue, we propose content-aware seam-carving (CSC), which incorporates semantic and spatial saliency priors. As illustrated by the blue region in Figure 2, the energy function of semantic seam-carving is formulated as follows:

$$Energy(I(x,y)) = |\frac{\partial}{\partial x}I(x,y)| + |\frac{\partial}{\partial y}I(x,y)| + S(x,y) \odot (1 - \frac{|x-x_0|}{W}), \tag{2}$$

where $\odot$ denotes element-wise multiplication, $S(x,y)$ represents the saliency value at position (x,y), x_0 is the x-coordinate of the saliency centroid achieved by averaging all saliency pixels, W is the width of the image. The saliency map can be obtained through a pretrained salient object detection network [17]. By enhancing the energy in important areas, the saliency prior $S(x,y)$ prevents key objects from deformation. The spatial prior $(1 - \frac{|x-x_0|}{W})$ further differentiates the importance within key regions, where the significance gradually decreases from the centroid towards the edges, thereby encouraging the model to prioritize seam removal from outer regions and retain key semantic elements for an object, ultimately avoiding distortion.

For convenience, we follow [1] to describe a vertical seam in an image as $s^x = \{s_i^x\}_{i=1}^W = \{(x(i), i)\}_{i=1}^W$, where $x(\cdot)$ is a mapping subject to $|x(i) - x(i-1)| \leq 1$. Given the energy function, we define the cost of a seam as $cost(s) = \sum_{i=1}^{W} Energy(s_i)$. The seam to be deleted is selected by minimizing the cost:

$$s^* = min_s \sum_{i=1}^{W} Energy(s_i). \tag{3}$$

Using dynamic programming, we can efficiently find the seams with the least energy. We set a tolerable saliency loss ratio λ to control the maximum loss of salient regions, which will be elaborated in Section 3.3.1. The maximun number of deleted seams is determined jointly by the saliency map and the tolerable saliency loss ratio λ. Specifically, the quantity of seams to be deleted, which intersect the saliency map, must not exceed the product of the saliency width W_s and the tolerable saliency loss ratio λ, whether the image reaches the target ratio. We further get a binary mask S where 0 represents the low energy pixel to delete and 1 is the pixel to be preserved. The initial retargeting results can be obtained by performing a dot product between the original image and the mask S and then concatenating the non-zero pixel regions.

3.3 Adaptive Repainting

The pixel-shift method inherently introduces pixel inconsistency, and bridging the resulting pixel gap poses a significant challenge. To address this issue, we introduce Adaptive Repainting (AR), a novel approach consisting of two primary components: the Adaptive Repainting Region Determination module (ARRD) and the Image-guided Repainting module (IR).

3.3.1 Adaptive Repainting Region Determination

The ARRD is designed to dynamically identify regions that require inpainting, which are characterized by inconsistencies among individual pixels. Additionally, ARRD determines the optimal repaint strategy (*i.e.*, inpaint or outpaint) and corresponding regions by comparing the current ratio with the target ratio.

To generate the inpainting mask, we identify pixels with a high number of deleted neighboring pixels in the content-aware seam carving (CSC) result as abrupt. As depicted by the green region in Figure 2, we employ a one-dimensional sliding window of length l on the mask map S of seam-carving to calculate the mean value within the window. This can be formalized as a one-dimensional convolution: $M = conv1d(S, K)/l$, where $conv1d$ is a one-dimension convolution operator, k is a one-dimensional convolution kernel of length l with all values equal to 1. We then binarize M into $\hat{M}$ using a threshold η, where 0 indicates areas to be inpainted and 1 denotes pixels to be preserved.

To generate the outpainting mask, we binarize the saliency map S into $\hat{S}$ using the mean value as the threshold. For each connected region in the saliency map, we compute its maximum width and then take the union of all these widths to obtain the saliency width W_s. This can be formalized as:

$$W_s = sum(Union(w_1, ..., w_H)), \tag{4}$$

where w_i is the *i*-th row of the binary saliency map $\hat{S}$, $Union(a, b)$ represents the union of two binary vectors a and b, and $sum(a)$ denotes the sum of all elements in the vector a. Given the target ratio r, we compare it with the target width W_t, which can be calculated as: $W_t = H * r$, where H is the height of original image. The final targeted width W_f can be determined as:

$$W_f = \begin{cases} W_s * (1 - \lambda), & W_s * (1 - \lambda) > W_t \\ W_t, & W_s * (1 - \lambda) \leq W_t \end{cases}, \tag{5}$$

where λ is the tolerable saliency loss ratio, set to 0.3 in our experiment. The final height H_f is then calculated as W_f/r, and we can determine if the image needs expanding by comparing H_f and H. The expanded height $(H_f - h)$ will be evenly distributed to the top and bottom of the image. The outpainting mask in this stage can be merged into the inpainting mask $\hat{M}$, hence the retargeting results can be obtained with a unified repainting process with $\hat{M}$.

3.3.2 Image-guided Repainting

As shown in Figure 3, to achieve repainting, a pretrained ControlNet [37], replicated from the Stable Diffusion (SD) [26] Unet, is parallelly combined with the SD model. This ControlNet (specifically,

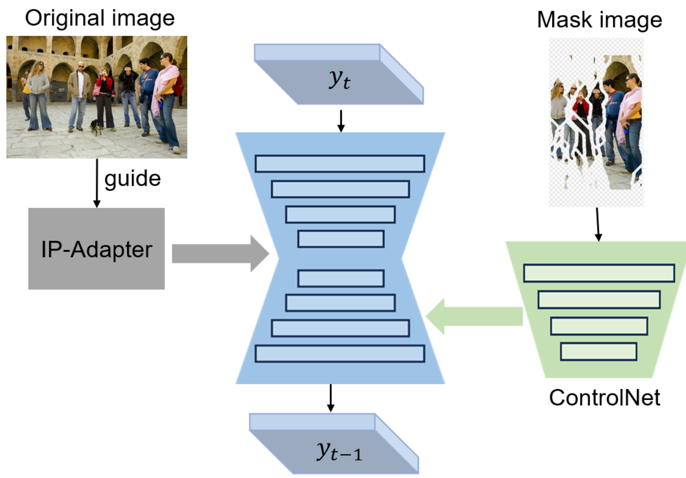


Figure 3: The architecture of image-guided repainting module.

the inpaint version) serves to introduce features associated with visible regions of the image to be repainted. To harness the guidance provided by the original image, we further introduce an IP-Adapter [36], which consists of a CLIP image encoder [25] and a lightweight adapter [12], to fuse image prompts with text prompts using decoupled cross-attention.

With the repainting mask $\hat{M}$ obtained in Section 3.3.1, we can formulate one reverse step in the diffusion process [19] to achieve repainting as follows:

$$y_{t-1} = \hat{M} \odot y_{t-1}^{known} + (1 - \hat{M}) \odot y_{t-1}^{unknown}, \tag{6}$$

where y_{t-1}^{known} is sampled with the unmasked pixels in the given image $\hat{M} \odot y_0$, while $y_{t-1}^{unknown}$ is sampled from the model with the previous iteration y_t.

4 Experiments

4.1 Dataset and Evaluation Metrics

We evaluate the proposed method on the public image retargeting datasets, RetargetMe [27], which contains 80 images from various scenes. According to the common sizes of prevalent electronic devices, we set the target aspect ratio for image retargeting as 16:9, 1:1, 4:3 and 9:16.

The metrics for image retargeting have remained undetermined and existing evaluation metrics [21, 18, 13] exhibit discrepancies with human perception, such as treating foreground and background equally. To intuitively evaluate the effectiveness of image retargeting methods, we propose **Saliency Discard Ratio** (SDR) to assess the semantic preservation. The SDR can be calculated as follows:

$$SDR = \frac{W_s^{ori} - W_s^{out}}{W_s^{ori}}, \tag{7}$$

where W_s^{ori} is the saliency width of the original image defined in equation 4 and W_s^{out} is the saliency width of the retargeted image.

User study metric. Given the subjective nature of retargeting results, we employ manual scoring as an additional evaluation method. Specifically, we invite 20 volunteers to rate the results on a scale from 0 to 3 across four aspects: content completeness, deformation, local smoothness, and aesthetics. These aspects are defined as follows: content completeness assesses whether key areas are cropped, deformation examines the degree of deformation within crucial areas, local smoothness evaluates the continuity of local regions in the image, and aesthetics evaluates the overall harmony and aesthetic appeal of the visual composition. A higher score indicates better performance.

4.2 Implement Details

Our method is implemented using Pytorch on a RTX 3090. The length of the sliding window in Section 3.3.1 is set to $l = 25$, and the threshold is set to $\eta = 15$. We utilize the VST model [17] for salient object detection in CSC. For the image-to-image repainting model in AR, we employ a composition of SD1.5[1], ControlNet-Inpainting[2] and IP-Adapter [36].

4.3 Compare with Other Retargeting Methods

We quantitatively evaluate the performance of our proposed model by comparing it with three other prevalent image retargeting methods, namely scaling, cropping, and seam-carving [1], using the objective metric 'SDR' and four subjective metrics across different aspect ratios.

Table 1: Comparison of SDR values with other retargeting methods on the RetargetMe dataset with different aspect ratios. Lower values indicate better semantic completeness. The best results are highlighted in **bold**.

Aspect Ratio	16/9	4/3	1/1	9/16
Scale	0.571	0.446	0.307	0.222
Crop	0.386	0.259	0.129	0.094
Seam-carving	0.490	0.367	0.242	0.161
Ours	**0.151**	**0.074**	**0.031**	**0.006**

Table 1 presents the performance of various methods on objective metric. As shown in the table, our method achieves a significant reduction in the loss of salient regions, primarily due to the incorporation of saliency priors.

Table 2: Subjective comparison with other retargeting methods in aspect ratio 16:9. ↑ indicates that larger are better. The best results are highlighted in **bold**.

Settings	Content completeness score ↑	Deformation score ↑	Local smoothness score ↑	Aesthetic score ↑	**Average score ↑**
Scale	**2.875**	0.975	1.878	1.153	1.720
Crop	1.295	**2.905**	**2.926**	2.355	2.370
Seam-carving	2.829	0.973	1.000	1.038	1.461
Ours	2.345	2.757	2.689	**2.538**	**2.582**

Table 2 presents the performance of different methods on four subjective evaluation metrics. As shown in the table, scaling and cropping exhibit two extremes, with scaling prioritizing content completeness and cropping prioritizing shape control. In contrast, our method receives high ratings across all four evaluation metrics. Notably, when compared to Table 1, scaling exhibits significant discrepancies between subjective and objective metrics in terms of key content preservation. We believe this is because the human eye has a natural interpolation ability compared to machines. Therefore, for scaling methods that uniformly delete pixels, subjective observers may not perceive strong content loss, even though objective metrics may indicate otherwise.

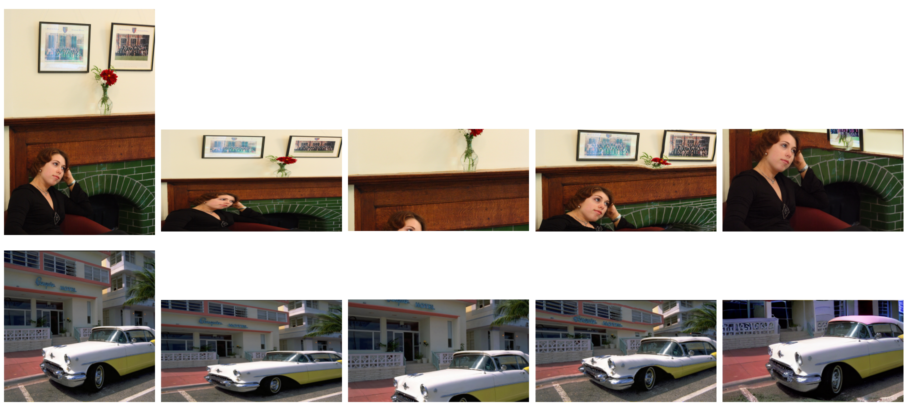
To qualitatively evaluate the performance of our proposed method, we visually compare our model with other 3 retargeting methods, including scaling, cropping and seam-carving [1] on different ratios. We conduct experiments with different ratios to provide an overall assessment of each method. Figure 4 and Figure 5 illustrate the comparison of retargeting results with two extreme target ratios respectively. We can intuitively observe that most traditional methods produce inferior results due to the lack of semantic information or the oversized salient areas. They struggle to balance the trade-off between preserving key content and preventing significant object deformation. In contrast, our proposed method effectively preserves the essential content and structure of foreground objects while simultaneously maintaining harmonious and consistent background.

[1] `https://huggingface.co/runwayml/stable-diffusion-v1-5`

[2] `https://huggingface.co/lllyasviel/control_v11p_sd15_inpaint`

Original image Scale Crop Seam-Carving Ours

Figure 4: Visual comparison to other retargeting methods on ratio 16:9



Original image Scale Crop Seam-Carving Ours

Figure 5: Visual comparison to other retargeting methods on ratio 9:16

4.4 Ablation Study

In this section, we comprehensively conduct ablation experiments to verify the effectiveness of each design in our proposed model on the popular aspect ratio 16:9.

Table 3: Ablation study of our retargeting methods on ratio 16:9.

Methods	SDR↓
Seam-carving	0.490
+CSC	0.190
+CSC+AR	**0.151**

Table 4: Comparison of background repainting and our adaptive repainting on ratio 16:9.

Methods	SDR↓
background repainting	0.190
adaptive repainting	**0.151**

Effectiveness of content-aware seam-carving. As shown in Table 3, content-aware seam-carving (denoted by '+CSC') significantly reduces the SDR, which means the salient objects are preserved much better. Besides, CSC can better preserve the structure of key objects, as evidenced by Figure

Figure 6: Visualization to demonstrate the effectiveness of each component in our method. '+CSC' denotes content-aware seam-carving in section 3.2, '+CSC+BR' means adopt background repainting based on CSC, '+CSC+AR' means adaptive repainting in section 3.3 based on CSC.

6. Different from the original seam-carving, the addition of the CSC module results in minimal deformation for the car. Also, the house with more complex patterns maintains its basic structure and avoids significant global deformation.

Effectiveness of adaptive repainting. The comparison between '+CSC' and '+CSC+AR' in Table 3 shows consistent improvement by the adaptive repainting module. As shown in Figure 6, AR adaptively identifies areas with abrupt pixels for repainting and adjusts the mask according to the target aspect ratio, leading to enhanced results.

Comparison between background repainting and adaptive repainting. To further validate the advantages of our proposed adaptive repainting method, we introduce the Background Repainting (BR) strategy for comparison. BR identifies the background based on saliency maps as the region for regeneration. Table 4 demonstrates the advantages of our AR method in preserving salient regions, which is supported by Figure 6. Specifically, BR is unable to address discontinuities in foreground pixels (see the car and the building in Figure 6), and the retargeting results are constrained by the ratio (the structure of car in the first row of Figure 6 due to extreme ratio). In contrast, our AR can identify all abrupt pixel regions and adapt well to extreme ratios.

4.5 Limitations

Constrained by the Stable Diffusion model, the inference speed of our method is relatively slow, averaging 7 seconds per image for retargeting. This may limit its real-time applicability in certain scenarios. Moreover, the repainting region generated by ARRD is not complete enough, as illustrated in the second row in Figure 6, certain distorted regions of the architectural structure remain unrepaired.

5 Conclusion

Our paper introduces a new image retargeting model called PrueRepaint. This model is content-aware and adaptive, allowing it to work with any target ratio. The content-aware seam-carving method protects important semantic regions, while the adaptive repainting module helps to maintain visual quality even after pixels are deleted. Through extensive experiments, we have demonstrated the effectiveness of our design and the advantages of using PrueRepaint for image retargeting.

References

[1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. In *ACMSIGGRAPH*, page 10–es, New York, NY, USA, 2007. Association for Computing Machinery.

[2] J. Canny. A computational approach to edge detection. *TPAMI*, (6):679–698, 1986.

[3] D. Cho, J. Park, T.-H. Oh, Y.-W. Tai, and I. So Kweon. Weakly-and self-supervised learning for content-aware deep image retargeting. In *ICCV*, pages 4558–4567, 2017.

[4] E. Dickman, P. Diefenbach, M. Burlick, and M. Stockton. Smart scaling: A hybrid deep-learning approach to content-aware image retargeting. In *ACMSIGGRAPH*, pages 1–2. 2023.

[5] W. Dong, N. Zhou, J.-C. Paul, and X. Zhang. Optimized image resizing using seam carving and scaling. *TOG*, 28(5):1–10, 2009.

[6] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification and scene analysis*, volume 3. Wiley New York, 1973.

[7] J. B. Dy, J. J. Virtusio, D. S. Tan, Y.-X. Lin, J. Ilao, Y.-Y. Chen, and K.-L. Hua. Mcgan: mask controlled generative adversarial network for image retargeting. *Neural. Comput. Appl*, 35(14):10497–10509, 2023.

[8] X. Fan, J. Lei, J. Liang, Y. Fang, X. Cao, and N. Ling. Unsupervised stereoscopic image retargeting via view synthesis and stereo cycle consistency losses. *Neurocomputing*, 447:161–171, 2021.

[9] X. Fan, Z. Zhang, L. Sun, B. Xiao, and T. S. Durrani. A comprehensive review of image retargeting. *Neurocomputing*, page 127416, 2024.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *COMMUN ACM*, 63(11):139–144, 2020.

[11] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *NeuriPS*, 33:6840–6851, 2020.

[12] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly. Parameter-efficient transfer learning for nlp. In *ICML*, pages 2790–2799. PMLR, 2019.

[13] C.-C. Hsu, C.-W. Lin, Y. Fang, and W. Lin. Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss. *IEEE J. Sel. Top. Signal Process*, 8(3):377–389, 2014.

[14] N. Kajiura, S. Kosugi, X. Wang, and T. Yamasaki. Self-play reinforcement learning for fast image retargeting. In *ACM Multimedia*, 2020.

[15] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998.

[16] J. Lin, T. Zhou, and Z. Chen. Deepir: A deep semantics driven framework for image retargeting. In *ICMEW*, pages 54–59. IEEE, 2019.

[17] N. Liu, N. Zhang, K. Wan, L. Shao, and J. Han. Visual saliency transformer. In *ICCV*, pages 4722–4732, 2021.

[18] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu. Image retargeting quality assessment. In *Computer Graphics Forum*, volume 30, pages 583–592. Wiley Online Library, 2011.

[19] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, pages 11461–11471, 2022.

[20] L. Ma, W. Lin, C. Deng, and K. N. Ngan. Image retargeting quality assessment: A study of subjective scores and objective metrics. *IEEE J. Sel. Top. Signal Process*, 6(6):626–639, 2012.

[21] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *TCSVT*, 11(6):703–715, 2001.

[22] Y. Mei, X. Guo, D. Sun, G. Pan, and J. Zhang. Deep supervised image retargeting. In *ICME*, pages 1–6. IEEE, 2021.

[23] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein. Unrolled generative adversarial networks. In *ICLR*, 2016.

[24] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023.

[25] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763. PMLR, 2021.

[26] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022.

[27] M. Rubinstein, D. Gutierrez, O. Sorkine-Hornung, and A. Shamir. A comparative study of image retargeting. *ACMSIGGRAPH*, 2010.

[28] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. *TOG*, 27(3):1–9, 2008.

[29] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen. Gaze-based interaction for semi-automatic photo cropping. In *SIGCHI*, pages 771–780, 2006.

[30] T. R. Shaham, T. Dekel, and T. Michaeli. Singan: Learning a generative model from a single natural image. In *ICCV*, pages 4570–4580, 2019.

[31] A. Shocher, S. Bagon, P. Isola, and M. Irani. Ingan: Capturing and retargeting the" dna" of a natural image. In *ICCV*, pages 4492–4501, 2019.

[32] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[33] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[34] D. Valdez-Balderas, O. Muraveynyk, and T. Smith. Fast hybrid image retargeting. In *ICIP*, pages 1849–1853. IEEE, 2021.

[35] D. Vaquero, M. Turk, K. Pulli, M. Tico, and N. Gelfand. A survey of image retargeting techniques. In *ADIP XXXIII*, volume 7798, pages 328–342. SPIE, 2010.

[36] H. Ye, J. Zhang, S. Liu, X. Han, and W. Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.

[37] L. Zhang, A. Rao, and M. Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, pages 3836–3847, 2023.

[38] M. Zhang, L. Zhang, Y. Sun, L. Feng, and W. Ma. Auto cropping for digital photographs. In *ICME*, pages 4–pp. IEEE, 2005.

[39] Y. Zhou, Z. Chen, and W. Li. Weakly supervised reinforced multi-operator image retargeting. *TCSVT*, 31(1):126–139, 2020.

A Appendix / supplemental material

A.1 Additional Comparative Experiments

Effectiveness of Space Prior. Considering the adverse impact of severe foreground distortions on aesthetic appeal, we introduce a spatial prior in Section 3.2 to differentiate the degrees of importance within the foreground region, where pixel importance decays from the centroid outwards. As depicted in Figure 7, the space prior allows the foreground region to preferentially lose pixels on both sides within a certain range to maintain the structure of important areas, which better conforms to human visual perception.

Original image Seam-carving +CSC w/o space prior +CSC w/ space prior

Figure 7: Visualization to demonstrate the effectiveness of space prior.

Comparison of different generation methods. We compare the differences among various generation methods, including full-image repainting, background repainting, and our adaptive repainting. The generated images are illustrated in Figure 8. Without the guidance of mask image, FR maintains coarse semantics but loses details. BR struggles to handle the issue of discontinuous foreground pixels (as evident in Figure 8 with the legs of the person and the pillars on both sides of the Taj Mahal). However, our AR can identify prominent areas of pixel discontinuity across the entire image for repainting.

A.2 Detail of Subjective Experiments

Regarding the subjective experiment, we provided brief training to the volunteers, covering an introduction to image retargeting, the differences between the four evaluation metrics, and the scoring range. Subsequently, we distributed questionnaires containing the results of 80 images from the retargetme dataset (reduced to 79 images after removing sensitive content) under four different methods. The questionnaire collection time was 2 to 3 hours, and we obtained 20 valid responses. We calculated the average scores for each evaluation metric. The specific description in the questionnaire is as follows: The left side shows the original image, and the right side shows the image after being

Figure 8: Comparison of different generation methods. '+CSC' denotes the content-aware seam-carving in section 3.2, 'FR' means full-image repainting, 'BR' represents the background repainting and 'AR' is our adaptive repainting.

retargeted to a 16:9 aspect ratio. Please rate each processed image from 0 to 3 (higher scores indicate better effectiveness). The compensation given to each volunteer was 500 RMB.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims that we propose a content-aware image retargeting method for any ratio reflect the scope, and the contribution is discussed in Abstract and Section 1 clearly.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of our work in the section 4.5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This paper do not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper elaborates on the model details comprehensively in Section 3, providing sufficient information for reproducibility.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: All experiments are conducted on publicly available datasets, and we plan to release the code soon for open access.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper carefully outlines the experimental details and hyperparameters in Section 4.2.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Our method employs pretrained models and maintains a consistent random seed; hence, there are no error bars.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper lists the computational resources is a single RTX 3090 in Section 4.2. The inference speed for retargeting is approximately 7 seconds per image.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: All experiments in this paper adhere to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper proposes a novel image retargeting model for any ratio, as discussed in Section 1, which has the potential to advance the field of image retargeting and contribute positively to the retargeting society. As far as we know, there is no negetive societal impact of the work performed.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: As far as we know, the paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We ensure that the assets we use are licensed.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We will provide the corresponding assets after the paper is accepted.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: The full text of instructions given to participants to subjectively assess the retargeting results and the compensation is listed in Appendix Section A.2.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: As far as we know, our work does not involve potential risks incurred by study participants. We obtain relative approvals to collect the evaluation results.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.