DEEP SUPERVISED IMAGE RETARGETING

Yijing Mei, Xiaojie Guo, Di Sun, Gang Pan and Jiawan Zhang*

College of Intelligence and Computing
Tianjin University, Tianjin 300350, China
pangang@tju.edu.cn

2021 IEEE International Conference on Multimedia and Expo (ICME) | 978-1-6654-3864-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICME51207.2021.9428129

ABSTRACT

Recent learning-based image retargeting methods have achieved significant improvement. However, two main issues remain in this challenging task: (i) it is difficult to build ground truth datasets for supervised learning; (ii) most methods are based on a certain operator, not suitable for various images with different target sizes. In this paper, for the first time, we address these issues by providing a deep supervised image retargeting solution. We introduce a new dataset[1] of $6,576$ pairs generated by multiple operators using Image Retargeting Quality Assessment (IRQA) algorithm. We then develop a mult-operator image retargeting model named MR-GAN, which learns the deformation process of retargeted images using multiple methods and conducts retargeting operations in feature space. Experimental results validate the effectiveness as well as its superiority against state-of-the-art alternatives of the proposed approach.

***Index Terms*—** Image retargeting, supervised, generative adversarial network, assessment-based dataset.

1. INTRODUCTION

Image retargeting aims to adapt an image to its best possible look that satisfies the target display device. It has been widely studied in the last decades with content-aware methods [1, 2, 3, 4, 5] as representatives. These approaches use low-level semantic features to perform retargeting operations in the image space, thus causing visual distortion or artifacts.

Recently, some deep learning based methods [6, 7, 8] show improved performance in image retargeting, as they can capture higher level semantic features of an original image. However, few studies have used supervised techniques because of the lack of image retargeting dataset for accurately training deep models in feature space. In addition, they heavily rely on calculated importance maps, which frequently suffer from inaccurate predictions and visual artifacts. Moreover, they perform retargeting in image space similarly with the traditional methods, hardly avoiding unnatural distortions in results. Therefore, it is necessary to construct a large-scale image retargeting dataset for accurately training deep models in feature space.

In addition, a certain retargeting operator may not perform well on all cases and all sizes. The Multi-operator [3] and Photo Squarization [6] start directly from the order of the original operators, expecting to seek the optimal combination for different operators and then execute directly in the pixel domain. But the execution unit of different operators for retargeting are not identical, so it is unlikely to incorporate retargeting methods with different execution units, and the results cannot be consistently executed across a larger range of images. Therefore, synthesizing multiple operators form dataset is considered as an effective solution.

This paper is, to the best of our knowledge, the first attempt to study the problem of image retargeting in a supervised way. One challenge is shared with all the supervised methods: the training process needs paired data as ground truth. But the way to produce the ground truth is uncertain and the method to evaluate retargeting results is usually subjective. In addition, it has two unique scientific challenges: (i) Image retargeting requires a mechanism to capture the similarities with different aspect ratios, and (ii) Collecting a large-scale dataset from different operators needs a lot of efforts.

Table 1. Multi IRQA-based retargeting datasets.

original dataset	resolution	total amount	training set	testing set
HKU-IS	224×224	$4,236$	$4,000$	236
COCO	300×300	$6,576$	$5,000$	$1,576$
Waterloo Exploration	300×300	$4,532$	$4,000$	532
AVA	300×300	$6,500$	$6,000$	$1,500$

We address the above challenges by contributing a large-scale dataset and developing a deep supervised approach for image retargeting. For the dataset, we introduce IRQA algorithm [9] to evaluate the retargeted images generated by different operators. The dataset collects various retargeting results from different operators into a group. Then, the results with the highest evaluation score in each group are regarded as the (pseudo) ground truth. For the network, we propose a multi-operator image retargeting network inspired by GANs, called MRGAN. MRGAN learns how the original image is converted into the "ground truth" in a complete end-

* Corresponding author. This work was supported by key research and development plan support program of Tianjin(20YFZCSN01080) and Tianjin transportation science and technology development project (2016A-02-01).

[1]https://github.com/TIReD2020/TIReD

978-1-6654-3864-3/21/$31.00 ©2021 IEEE

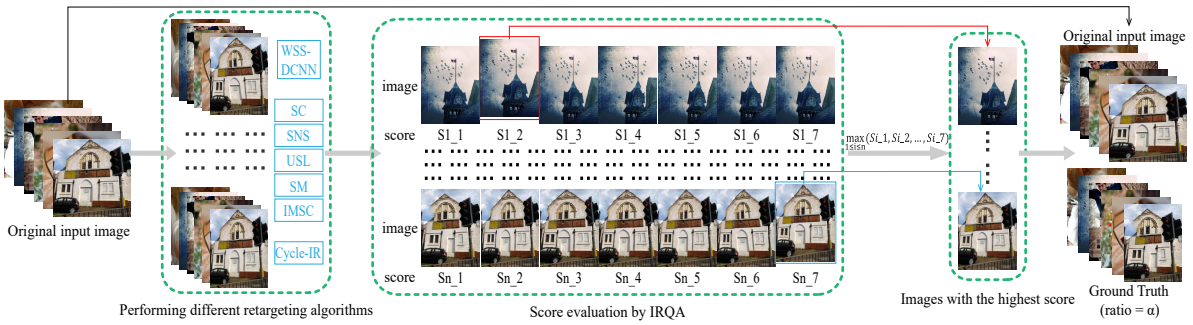


Fig. 1. Overview of Data Collection. We use the collected original image as input. Then we use the IRQA algorithm in [14] to generate the retargeted image as the ground truth. (The width scaling ratio sets 0.75)

to-end manner. Because MRGAN is driven by dataset which is a combination of multiple operators, it is not limited by the execution unit and can be more widely used. In addition, MRGAN performs retargeting in feature space, it is possible to repair details through image reconstruction while learning the overall deformation of target images, thus improving the visual quality.

The contributions are as follows: (1) This study presents a deep supervised image retargeting solution that consolidates information across multiple operators to evaluate the retargeted images. A new dataset for the retargeting tasks is constructed. (2) We propose a multi-operator image retargeting network (MRGAN), which can perform retargeting operations in feature space. (3) In-depth experiments are conducted to show that our method is potentially able to restore unnatural distortions of the retargeted image through image reconstruction.

2. IRQA-BASED RETARGETING DATASET

One of the key barriers to deep supervised retargeting research is lack of benchmark datasets. Most existing retargeting datasets, such as RetargetMe [10], NRID [11], and CUHK [12], are created by various images and the results of previous retargeting operators on these images. Therefore, they do not actually have ideal target images, and are too small to support deep network training.

We contribute a dedicated dataset for image retargeting. There are $6,576$ image pairs in total. Each pair consists of an original image and a corresponding target image which is selected from results of multiple operators using IRQA algorithm. The training set takes $5,000$ pairs, while the rest $1,576$ pairs go to the testing set.

When collecting the images, we selected the COCO dataset [13] according to the sufficient amount of data, the diversity and complexity of the images. Our goal is to achieve the retargeting by learning target images using a GAN network. The focus and difficulty of this process is how to select target images. In the field of image retargeting, there are many mature IRQA algorithms [14, 10, 15, 9]. To the best of our knowledge, we are the first to work with IRQA in data manufacture during retargeting process. We create the dataset through the following four steps (as shown in Figure 1):

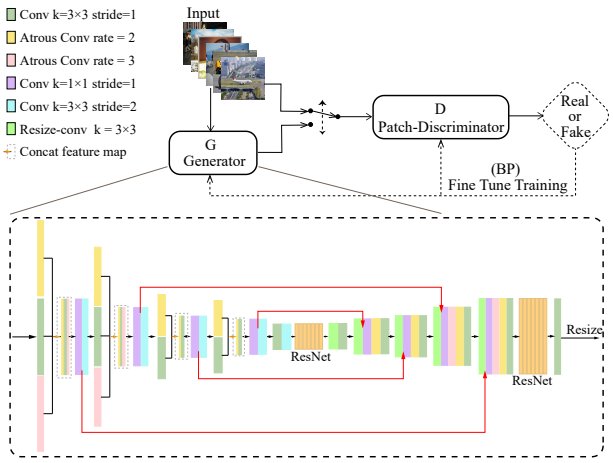


Fig. 2. Overview of MRGAN. It consists of a generator network and a patch discriminator network. The red lines in generator indicate skip-connections, and k is the kernel size.

(1) Determining the input. We normalize the size of the collected $6,576$ original images to 224×224, which are then used as input for all subsequent operations.

(2) Performing retargeting. We implement certain previous representative retargeting algorithms, including: uniform scaling (USL), seam-carving (SC) [1], improved seam-carving (ISC) [2], scale-and-stretch (SNS) [5], shift map (SM) [4], Cycle-IR [8], WSSDCNN [7]. We employ seven retargeting operators to perform the identical retargeting on the same original input image. The input image and its retargeted versions (8 images in total) form a group.

(3) Assessing scores for retargeted images. We take the IRQA algorithm as the evaluation criteria and use the method in [9] to measure the retargeted image based on the original input image in each group.

(4) Retaining the highest score image. We treat the retargeted image with the highest score in the group as the ground truth, which is used for training.

Moreover, in order to better verify that the proposed method can achieve preferable results on various and multi-

resolution images, this paper also constructs several retargeting datasets at different resolutions based on a variety of different original datasets using the above method. Among them, the resolutions are 224×224 and 300×300 respectively. Due to the limitation of WSSDCNN model on image input size, we do not use this method when generating the dataset with a resolution of 300×300. And the details of remaining established datasets are shown in the Table 1.

3. METHODOLOGY

3.1. MRGAN Model

Our model(as shown in Figure 2) is inspired by Generative Adversarial Networks (GANs) [16]. It contains two parts, a generator G and a discriminator D.

Generator. The generator G follows the encoder-decoder fashion, and uses the structure of U-Net [17]. The architecture of the generator G is shown in Figure 1. The encoder part has 20 convolutions layers, which are divided into 5 convolutional blocks. From shallow to deep, the convolution blocks contain 5, 5, 4, 4 and 2 convolution layers respectively, including convolution layers with the 3×3 kernel, dilated convolutions [18] (it could be zero), and a 1×1 convolution layer. The first four blocks all end up with a convolution layer with a stride of 2. Among them, the network concatenates feature maps of the first convolution layer and the dilated convolution layer(s) in the convolution block with the dilated convolution, and sends it to the next layer. The encoder component is followed by ResNet module [19] that contains six residual blocks. The decoder is completely symmetrical to the encoder. In order to avoid checkerboard-like artifacts, we use an alternative to regularize the deconvolution, *i.e.* resize-convolution layer[2].Therefore, the deconvolution layer, which corresponds to the convolution layer with a stride of 2 in the encoder, employs the resize-convolution layer.

For the skip-connections between the encoder and decoder, we connect the output of the penultimate convolution layer for each convolution block in the encoder with the feature map, which is obtained after the resize-convolution layer of the corresponding deconvolution block in the decoder. After the encoder-decoder, the network further employs 6 residual blocks to learn/recall details preferably, and then it generates retargeted image through a regular 3×3 convolution. Since the size of the original image and the retargeted image are inconsistent, we interpolate the ground truth to the same size as original image and use it as the target retargeted image for the G to learn. The size of retargeted image initially generated by G is equal to that of the original image, so the image is required to be uniformly scaled after the decoder to obtain the width-reduced retargeted image.

Discriminator. It uses the patch discriminator [20] with full convolution for discriminator D. We can see that the difference between the generated image and the original image for image retargeting is relatively small, and patches in patch discriminator have preferable local characteristics. By comparing the differences of local features between patches, the local differences and dynamics of retargeting operation are perceived more accurately. During the training, the generated image and ground truth are sent into D together, and an $N \times N$ matrix is finally output, where N is the size of patch, and the value of matrix represents the discrimination result of patch.

[2]http://distill.pub/2016/deconv-checkerboard

3.2. Loss Design

Content Loss. In order to better capture the deformation characteristics of retargeting and improve the mapping accuracy from the input to the retargeted image, we design a content loss $\mathcal{L}_{con}$ for the skip-connections part of G:

$$\mathcal{L}_{con} = \alpha \cdot \frac{1}{n-1} \sum_{i=2}^{n} \parallel F(i)_{Ori} - \Phi(i)_{GT} \parallel_1, \quad (1)$$

where i represents the i-th connection from outer to the inner side in Figure 1, and n is the number of connections experienced in the network. $F(\cdot)_{Ori}$ expresses the feature map output by the encoder when going through the i-th connection. $\Phi(\cdot)_{GT}$ donotes the feature map of interpolated ground truth after the same number of convolution blocks in the pre-trained *VGG*19, and it is also the target feature map that G needs to learn. $\|\cdot\|_1$ is the $\mathcal{L}_1$ norm. In addition, α is used to provide the contribution weight for this item to update the parameters.

Total Variation Loss. To enhance the local smoothness and eliminate artifacts in generated images, we employ a total variation loss $\mathcal{L}_{tv}$ as:

$$\mathcal{L}_{tv} = \frac{1}{HWC} \parallel \nabla G(x) - \nabla x_{gt} \parallel_1, \quad (2)$$

where H is height, W is width, C is the number of channels, and ∇ is the gradient of the image. To enforce the edge of the generated image, we further define the $\mathcal{L}_{m\_tv}$ based on the total variation loss:

$$\mathcal{L}_{m\_tv} = \frac{1}{HWC} \parallel M_{x_{gt}} \cdot (\nabla G(x) - \nabla x_{gt}) \parallel_1, \quad (3)$$

where x_{gt} is the corresponding width-reduced image in the dataset of ground truth, and x is the original input image. M represents the Canny edge map.

Adversarial Loss. The GANs need to train both the generator and the discriminator simultaneously to make them adversarial. We use the standard adversarial loss $\mathcal{L}_{MRGAN}(G, D)\&$ as follows:

$$\begin{aligned} \mathcal{L}_{MRGAN}(G, D) \; = \; & \mathbb{E}_{x_{gt} \sim p_{data}(x_{gt})}[\log D(x_{gt})] \\ & + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D(G(x)))], \end{aligned} \quad (4)$$

Other Losses. Some extra losses are employed for the training stability. First, we use the $\mathcal{L}_1$ reconstruction loss [20]

to emphasize the matching of each corresponding pixel between the generated image and the real image.

$$\mathcal{L}_1 = \| x - G(x) \|_1, \tag{5}$$

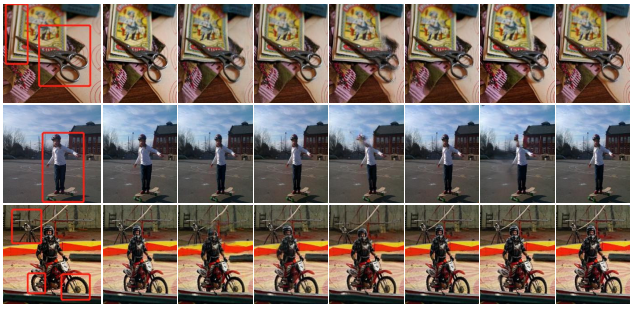
A perceptual loss $\mathcal{L}_P$ is also adopted to encourage similar features between the generated images and the real images. Images trained through this loss generally contain more high-frequency information. The loss is given as:

$$\mathcal{L}_P = \sum_{l=L-2}^{L-1} \beta_l \cdot \| \Phi_l(x_{gt}) - \Phi_l(G(x)) \|_2^2, \tag{6}$$

where L denotes the numbers of convolution blocks in the pre-trained $VGG19$. Φ_l is the function for extracting the feature map after the l-th convolution block in $VGG19$. $\| \cdot \|_2$ is the Euclidean norm. β is used to provide the contribution weight for each item to the total loss. To maintain the structural similarity more accurately, we further introduce a structural similarity loss $\mathcal{L}_{SSIM}$ as:

$$\mathcal{L}_{SSIM} = \frac{1}{B} \sum_{b=1}^{B} [1 - SSIM(G(x), x_{gt})], \tag{7}$$

where B is the size of batch size, and $SSIM$ is calculated by $G(x)$ and the ground truth.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3. Ablation study. (a) Original input images. (b) Results of our method. (c)-(h) Retargeted results that training MRGAN without $\mathcal{L}_{con}$, $\mathcal{L}_1$, $\mathcal{L}_{tv}$, $\mathcal{L}_{m\_tv}$, $\mathcal{L}_P$ and $\mathcal{L}_{SSIM}$, respectively.

Total Objective. The overall loss G^* is defined as follow, where λs are importance coefficients:

$$\begin{aligned} G^* = \arg\min_G \max_D L_{MRGAN}(G, D) + \lambda_P \mathcal{L}_P + \lambda_{\mathcal{L}_1} \mathcal{L}_1 \\ + \lambda_{tv} \mathcal{L}_{tv} + \lambda_{m\_tv} \mathcal{L}_{m\_tv} + \lambda_S \mathcal{L}_{SSIM}. \end{aligned} \tag{8}$$

4. EXPERIMENT

In training, we set the value of α in $\mathcal{L}_{con}$ to 0.1. We also set the value of β_{L-1} in $\mathcal{L}_P$ to 2 and β_{L-2} to 1. And in formula 8, the value of $\lambda_{\mathcal{L}_1}$, λ_P, λ_{tv}, λ_{m_tv}, and λ_S is 1000, 10, 0.01, 0.003, 70, respectively. Each convolutional layer contains the activation layer and a batch normalization after convolution. For the activation function, the encoder uses LReLU and the decoder employs ReLU. In the experiment, our model perform a complete end-to-end training using the Adam optimizer. The batch size is set to 4. We train the model on an NVIDIA Titan V GPU.

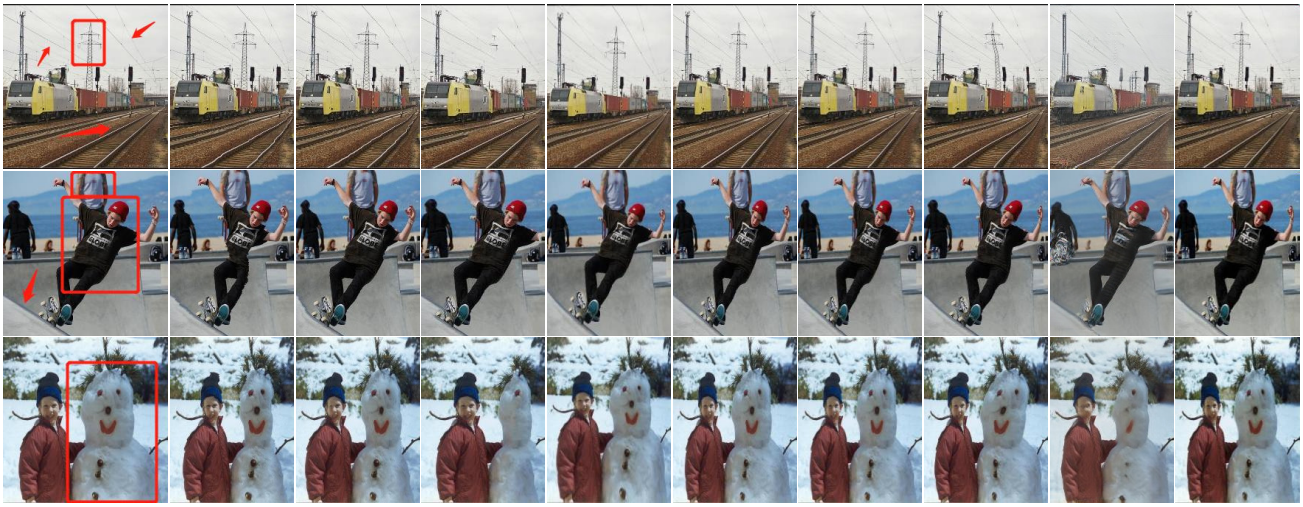
Table 2. Retargeting results in terms of PSNR, SSIM, FSIM and VIF for ablation study. The values in bold means the best performances.

Method	no-$\mathcal{L}_{con}$	no-$\mathcal{L}_P$	no-$\mathcal{L}_{tv}$	no-$\mathcal{L}_{m\_tv}$	no-$\mathcal{L}_1$	no-$\mathcal{L}_{SSIM}$	Ours
PSNR	21.8165	21.6348	21.6259	21.8882	21.7676	21.6903	**21.9360**
SSIM	0.6280	0.6257	0.6295	0.6355	0.6355	0.6356	**0.6375**
FSIM	0.8073	0.8055	0.8077	0.8118	0.8115	0.8114	**0.8125**
VIF	0.2726	0.2707	0.2704	0.2838	0.2841	0.2812	**0.2886**

Ablation Study. We find that these loss functions can all improve the effect of the model, but the action direction is slightly different. For example: $\mathcal{L}_{con}$ and $\mathcal{L}_P$ improve the correctness of the semantics in the retargeted image. As shown in Figure 3, semantic information in the last two cases are better preserved. And the design of $\mathcal{L}_{tv}$ and $\mathcal{L}_1$ improve the visual reality of the results(the first row in Figure 3). At the same time, from the first and last rows, we can see that $\mathcal{L}_{SSIM}$ provides a great contribution to the maintenance of straight lines and object shapes in the image. Finally, $\mathcal{L}_{m\_tv}$ enables the model to better maintain the original contour in the original image, as shown in Figure 3, the roundness of the wheels are better remained. In addition, PSNR, SSIM [21], FSIM [22], and VIF [23] are all commonly used image evaluation. The larger their value, the better the image quality. It can be seen from Table 2 that only when all the loss functions are added to the model, the evaluation result is the best. That is, each loss function has a certain optimization effect for MRGAN in different directions.

Experiments with Different Datasets and Multi Resolutions. In this part, we first conduct experiments on different datasets, and the results are shown in Figure 6. Because many different datasets contain many a variety of images and these images have different resolutions, the results in this part are very convincing. In addition, the proposed model is built based on fully convolutional layers, so MRGAN can handle input images of any size. Figure 6 shows results for multi resolution images. It can be seen that once MRGAN is trained, it is able to well preserve the important areas of input images (such as people, animals, and cars) and the overall structure (straight line in the figure) over a large resolution span (224×224 to 500×500). Therefore, this visual example demonstrates the powerful retargeting ability of MRGAN for images of multi resolutions.

Comparison with Previous Methods. We compare with the previous methods on the proposed dataset and benchmark dataset RetargetMe. Figure 4 depicts several visual examples, where the first two rows are from the proposed dataset



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

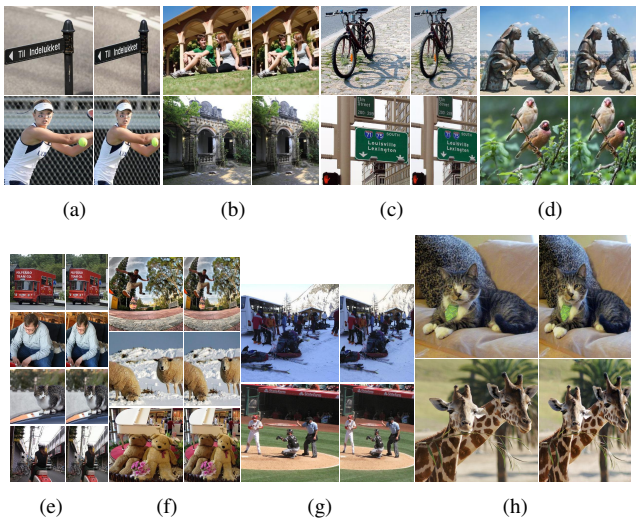
Fig. 4. Visual comparison with the classic methods implemented in the paper. (a) Original input images. (b)-(j) Retargeted images from SC, ISC, SM, SNS, USL, Cycle-IR and WSSDCNN, InGAN, and our method, respectively.

while the last two rows from the RetargetMe dataset. It can be seen that our approach is a comprehensive learning of different methods, so it is more suitable for different types of images. Besides, our method considers semantic information and makes semantic expression more accurate. Because our method is to learn the deformation of ground truth, and then reconstruct the result from the original image. So our results have superior visual perception than ground truth. As shown in Figure 5, our results repair the excessive distortions and fill in the absence of important areas in ground truth.

(a) Original (b) GT (c) Ours (a) Original (b) GT (c) Ours

Fig. 5. Repair effect of the proposed method.

Quantitative Evaluation. To ensure fairness and rationality, we use two image quality evaluation standards Natural Image Quality Evaluator[3] (NIQE) [24], Blind/Referenceless Image Spatial QUality Evaluator[4](BRISQU) [25] and Novel Blind Image Quality Assessment (NBIQA) [26] that do not require ground truth to verify our method. The NIQE has a good consistency with the subjective quality evaluation of the human eye. The BRISQU evaluates image from a spatial perspective, and the NBIQA considers features from both spatial domain and transform domain. The experiments are conducted on two datasets including the proposed dataset and RetargetMe dataset. Table 3 presents the quantitative results of ours and 7 state-of-the-art methods, including SC, ISC, SM, SNS, USL, Cycle-IR and WSSDCNN. For all these algorithms, higher value represents a lower quality. As shown, our model achieves the best performance on all datasets in all terms.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 6. Results for different datasets and multi resolutions. For each image pair, the left is original image, and the right is retargeted image. (a)-(d) are from COCO, Waterloo Exploration, AVA and HKU-IS, respectively. And the original image resolutions of (e)-(h) are 224×224, 300×300, 400×400, and 500×500, separately.

5. CONCLUSION

In this paper, we introduced a novel task of deep supervised image retargeting. This task is more challenging than the well studied non-supervised retargeting task, but is also more useful for multifarious images. A new IRQA-based dataset which contains results of various retargeing methods were introduced to stimulate the research in this direction, so it is theoretically possible to gather the advantages of various retargeting methods. We further presented a deep model to gen-

[3]http://live.ece.utexas.edu/research/quality/niqe.zip
[4]http://live.ece.utexas.edu/research/quality/BRISQUErelease.zip

erate retargeted images using multi-operators in a complete end-to-end manner. The proposed method is capable of producing visually pleasing target images. We provided in-depth comparisons with existing approaches on both our and RetargetMe datasets and demonstrated the superior performance of our approach.

Table 3. Retargeting results in terms of NIQE, BRISQU and NBIQA on the new dataset and RetargetMe(RM). The values in bold indicate the best performances of different methods.

Method	NIQE	BRISQUE	NBIQA	NIQE (RM)	BRISQUE (RM)	NBIQA (RM)
SC	6.1959	25.306	24.085	7.2971	29.540	24.950
ISC	6.2171	23.564	23.508	7.4080	28.668	24.525
SM	6.1496	23.398	24.441	6.8255	25.925	25.951
SNS	5.7299	20.302	24.393	5.9060	17.364	20.899
USL	6.0372	22.146	24.007	6.6481	23.827	23.478
Cycle-IR	5.8608	22.562	23.211	6.1486	21.008	20.953
WSSDCNN	5.9312	22.014	23.300	6.4185	23.161	23.216
Ours	**5.5026**	**18.158**	**21.560**	**5.7186**	**16.535**	**19.876**

6. REFERENCES

[1] Shai Avidan and Ariel Shamir, "Seam carving for content-aware image resizing," *ToG*, July 2007.

[2] Michael Rubinstein, Ariel Shamir, and Shai Avidan, "Improved seam carving for video retargeting," *ToG*, pp. 16:1–16:9, Aug. 2008.

[3] Michael Rubinstein, Ariel Shamir, and Shai Avidan, "Multi-operator media retargeting," *ToG*, p. 23, 2009.

[4] Yael Pritch, Eitam Kav-Venaki, and Shmuel Peleg, "Shift-map image editing," in *ICCV*. IEEE, 2009, pp. 151–158.

[5] Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee, "Optimized scale-and-stretch for image resizing," in *ToG*. ACM, 2008, p. 118.

[6] Yu Song, Fan Tang, Weiming Dong, Xiaopeng Zhang, Oliver Deussen, Tong-Yee Lee, et al., "Photo squarization by deep multi-operator retargeting.," in *ACM Multimedia*, 2018, pp. 1047–1055.

[7] Donghyeon Cho, Jinsun Park, Tae-Hyun Oh, Yu-Wing Tai, and In So Kweon, "Weakly-and self-supervised learning for content-aware deep image retargeting," in *ICCV*, 2017, pp. 4558–4567.

[8] Weimin Tan, Bo Yan, Chumin Lin, and Xuejing Niu, "Cycle-ir: Deep cyclic image retargeting," *arXiv:1905.03556*, 2019.

[9] Feng Shao, Zhenqi Fu, Qiuping Jiang, Gangyi Jiang, and Yo-Sung Ho, "Transformation-aware similarity measurement for image retargeting quality assessment via bidirectional rewarping," *TSMC*, 2019.

[10] Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir, "A comparative study of image retargeting," in *ToG*. ACM, 2010, p. 160.

[11] Chih-Chung Hsu, Chia-Wen Lin, Yuming Fang, and Weisi Lin, "Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss," *JSTSP*, pp. 377–389, 2014.

[12] Lin Ma, Weisi Lin, Chenwei Deng, and King Ngi Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *JSTSP*, pp. 626–639, 2012.

[13] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *ECCV*. Springer, 2014, pp. 740–755.

[14] Susana Castillo, Tilke Judd, and Diego Gutierrez, "Using eye-tracking to assess different image retargeting methods," in *APGV*. ACM, 2011, pp. 7–14.

[15] Lin Ma, Weisi Lin, Chenwei Deng, and King N Ngan, "Study of subjective and objective quality assessment of retargeted images," in *ISCAS*. IEEE, 2012, pp. 2677–2680.

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *NeurIPS*, 2014, pp. 2672–2680.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[18] Fisher Yu and Vladlen Koltun, "Multi-scale context aggregation by dilated convolutions," in *ICLR*, 2016.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[20] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134.

[21] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *TIP*, vol. 13, no. 4, pp. 600–612, 2004.

[22] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "Fsim: A feature similarity index for image quality assessment," *TIP*, vol. 20, no. 8, pp. 2378–2386, 2011.

[23] Hamid R Sheikh and Alan C Bovik, "Image information and visual quality," *TIP*, vol. 15, no. 2, pp. 430–444, 2006.

[24] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *SPL*, pp. 209–212, 2012.

[25] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *TIP*, vol. 21, no. 12, pp. 4695–4708, 2012.

[26] Fu-Zhao Ou, Yuan-Gen Wang, and Guopu Zhu, "A novel blind image quality assessment method based on refined natural scene statistics," in *ICIP*. IEEE, 2019, pp. 1004–1008.