IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 10, 2019, accepted June 25, 2019, date of publication July 5, 2019, date of current version July 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927032

A Content-Aware Image Retargeting Quality Assessment Method Using Foreground and Global Measurement

YUWEI LI, LIHUA GUO, (Member, IEEE), AND LIANWEN JIN, (Member, IEEE)
School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China

Corresponding author: Lihua Guo (guolihua@scut.edu.cn)

This work was supported in part by the Guangzhou Science and Technology Program key projects under Grant 201707010141 and Grant 201704020134, and in part by the GD-NSF under Grant 2017A030312006.

ABSTRACT Image retargeting methods aim to minimize the perceptual loss while changing sizes and aspect ratios of images. Since optimal retargeting methods for different images are generally not the same, the image retargeting quality assessment (IRQA) becomes a meaningful task. This paper proposes a content-aware image retargeting quality assessment method using foreground and global measurement to achieve better performance. In our proposed method, images are first divided into two categories according to the foreground object detection result, and then different corresponding measurements are designed for them. For those with obvious foreground object, both foreground and global measurement are applied. For others, only global measurement is conducted. Foreground measurement includes two complementary features: the high-level semantic similarity feature and the low-level size ratio feature. Global measurement includes another two features: an improved aspect ratio similarity (ARS) feature and edge group similarity (EGS) feature. Two public databases, i.e., the RetargetMe and CUHK, have been evaluated, and experimental results demonstrate that our method is quite effective, and it also provides state-of-the-art performance in the IRQA.[a]

INDEX TERMS Image retargeting quality assessment, image quality assessment, foreground object quality measurement, semantic similarity feature.

I. INTRODUCTION

With the rapid development of mobile devices, image retargeting has become an urgent demand. Due to the fixed size and aspect ratio of mobile device, perceptual loss of image retargeting is almost unavoidable. Conventional retargeting methods, such as manual cropping (CR) and linear scaling (SCL), lead to content loss and structure loss respectively. Meanwhile, some content-aware retargeting methods [1]–[8] have been proposed in recent years which reduce overall perceptual loss by preserving more important regions while sacrificing less import regions.

Retargeting methods can be roughly divided into two categories: discrete methods and continuous methods. Discrete methods, e.g. seam-carving (SC) [1] and shift-map (SM) [6], directly remove or insert pixels in less important regions. These operators sometimes cause content loss and lead to discontinuity or artifacts. Continuous methods, e.g. steaming video (SV) [2], scale-and-stretch (SNS) [3] and non-homogeneous warping (WARP) [4], generate a sub-pixel level mapping for generating the retargeted image. Although content loss is reduced to a very low level by continuous methods, structure loss and shape distortion become two common issues. A particular retargeting operator cannot always produce the retargeted images with best perceptual quality. Therefore, it is a significant work to develop an effective image retargeting quality assessment method to help choose the best quality retargeted image and to improve retargeting methods.

In recent years, some works [11]–[13] show promising performance on IRQA problem. Basically, they partition the image into blocks, measure the quality of each block with a certain designed metric, and adopt saliency maps to produce a global quality score by summating up the saliency weighted

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Zhang.

[a]Our code are available at https://github.com/SCUT-ML-GUO/IRQA.

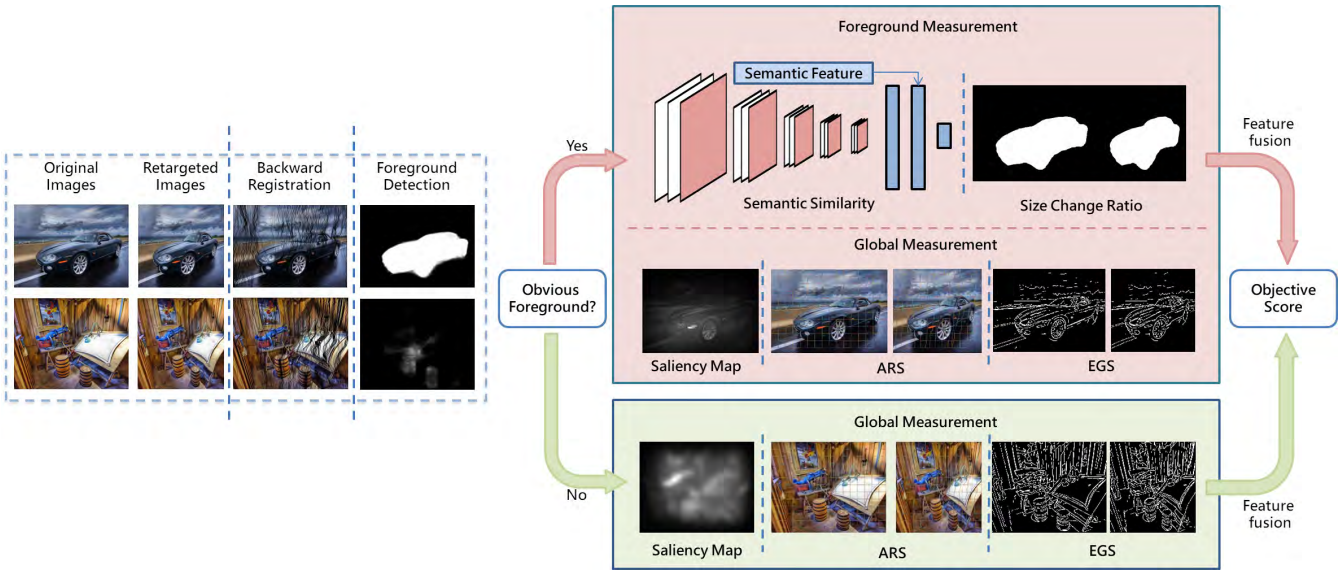


FIGURE 1. **The overall block diagram of our proposed method. Backward registration [9] and foreground object detection [10] are performed first. We judge image categories by the foreground object detection map, and thus decide whether we should conduct the foreground measurement besides the global measurement.**

scores of all the blocks. These low-level block-wise metrics mainly focus on structure loss and distortion, but they hardly involve semantic analysis of image content. IRQA is a kind of image aesthetic quality assessment, and it highly depends on the semantic component of images [14]. Therefore, it is significant to analyze the semantic component of images when evaluating image retargeting quality. Besides, existing methods neglect the inherent differences within different attributes of images, and they attempt to measure the perceptual quality of all images in a same evaluation standard.

From our observation, for images with obvious foreground object, we believe that individual analysis on foreground object is quite essential in IRQA. The main reasons are as follows: 1) the basic purpose of image retargeting is to protect the quality of salient contents, and thus the quality of foreground object is crucial to the overall quality. 2) the semantic analysis of image contents is complex. If we only measure the quality of the most important semantic component, i.e. the foreground object, then the analysis can be simplified while it remains effective. 3) the tolerance level of visual distortions in foreground is lower than that in background. Therefore, we define semantic similarity as a semantic loss and calculate the size change of foreground object to represent the perceptual loss of foreground object. While for images without obvious foreground object, we believe that the perceptual quality is mostly depended on the change of global structure, and thus we apply a global measurement for them.

Based on the aforementioned consideration, we propose a content-aware image retargeting assessment method in this paper, which is a IRQA framework by respectively designing the most suitable assessing measurement for two attributes of images, i.e. images with and without obvious foreground object. Our framework is the first one to design different measurement for different attributes of images. The overall framework is shown in Fig.1. First, the backward registration [9] is used to estimate the pixel-level correspondence between the original and retargeted images. Then, we divide images into two categories according to the foreground object detection result, and design different corresponding measurements for them. For those with obvious foreground object, we apply both foreground and global measurement. For others, only global measurement is conducted. Foreground measurement includes two complementary features of different levels: the high-level semantic similarity feature and low-level size ratio feature. The high-level semantic similarity feature is extracted to measure the similarity of foreground object output from a pre-trained neural network between original images and retargeted images. The low-level size ratio feature is calculated as the size change ratio of foreground object between original images and retargeted images. Global measurement includes another two features: an improved aspect ratio similarity (ARS) feature and edge group similarity (EGS) feature in [12]. For images with obvious foreground object, the quality of foreground regions plays a much more important role than background regions, while for images without obvious foreground object, the gap of importance between foreground quality and background quality is much smaller and thus we should take consideration about the global quality. Therefore, in our method, the saliency model that we adopt for images with foreground object or not are different. Finally, a learned model is used to predict the perceptual quality of retargeted images. Experiments on feature analysis in the section V demonstrate that the foreground measurement is very effective to evaluate the overall quality.

Although it is already a widely adopted approach to encode semantic components with CNN, and ARS and EGS have

been used in [12], our main contribution is that we establish a framework, in which we respectively design the most suitable assessing measurement for two attributes of images. Moreover, when extracting the semantic feature by CNN, our method focus on the foreground object and keep the original aspect ratio to maintain the key information in the preprocessing process. Therefore, this kind of feature is a more powerful representation of semantic content.

The paper is organized as follows. Section II introduces some related works on IRQA. In Section III, we describe technical details of our foreground measurement. Section IV introduces global measurement. Experimental results and analysis are given in section V. Finally, the conclusion is drawn in section VI.

II. RELATED WORKS

In recent years, many works [15]–[19] paid more attention in certain types of image quality assessment (IQA). Different from IQA, original images and retargeted images in IRQA were not in the same resolution. Therefore, structural similarity index (SSIM) [20] and peak-signal-to-noise ratio (PSNR) [21], which were widely used as metrics in IQA, were not suitable to measure the quality of retargeted images in IRQA.

Early image distance metrics, e.g. Edge Histogram (EH) [22], Color Layout (CL) [23], Bi-Directional Similarity (BDS) [24], SIFT-flow [9] and Earth mover's distance [25], have shown poor performance when predicting the image retargeting quality in the comparative study [26]. Recent studies took advantage of SIFT flow [9] to attain the alignment between original and retargeted images, making further analysis become available. Fang *et al.* [27] proposed IR-SSIM which applied SSIM between the matched local patches of original and retargeted images. Liang *et al.* [28] creatively considered aesthetics and symmetry measurement and achieved well correlated prediction with the subjective assessment. In [29], seven elaborate metrics were designed and all the features are fused using a General Regression Neural Network(GRNN). Jiang *et al.* [30] focused on learning a sparse representation of an image that contains distortion sensitive features. Oliveira *et al.* [31] measured the loss of relevant content and visual artifacts created in retargeted images in a bi-directional approach. Zhang *et al.* [32] analyzed in three levels including region-level, patch-level and pixel-level and made effort in detecting deformation inconsistency. Chen *et al.* [33] take advantage of gabor filters to extract log-Gabor statistical features and considered global structure distortion as well as salient area loss. Zhang *et al.* [11] proposed a simple but effective metric called aspect ratio similarity (ARS), which separated the original image into squares and took a consideration of the aspect ratio similarity between the original square and the corresponding patch in the retargeted image. In [12], two other measurements including edge group similarity (EGS) and face block similarity (FBS) have been complemented to improve the performance. Fu *et al.* [13] extracted texture feature and semantic feature using outputs of different layers in VGG16 besides some traditional hand-craft features.

The above methods mostly focused on low-level features and lack consideration in terms of semantic information. Besides, they only considered images as a whole and hardly involved specially designed analysis on foreground object. Two recent works [12], [13] did involve some attempt in either semantic analysis or individual analysis on foreground object. The semantics similarity measurement proposed in [13] uses a pre-trained convolutional neural network to encode the semantic components information of an image into a feature vector (denoted as semantic feature vector). Semantic component loss of a retargeted image is then measured by using an empirically designed similarity metric to evaluate the similarity between the semantic feature vectors of retargeted image and original image. However, they directly reshape the image to the input size of network, which causes two major problems: 1) Aspect ratio is a key factor of perceptual quality but the original aspect ratio information is completely lost due to the resizing operation; 2) All the semantic components of the whole image are encoded into one semantic feature vector in combination, which is a weak representation of semantic information. Face block similarity proposed in [12] has achieved a promising performance improvement in images with faces by specially measuring ARS of face blocks. However, the considered semantic component is simplified to one certain types of foreground object, i.e. human faces. Besides, semantic information requires high-level analysis but they only adopt one low-level feature, i.e. ARS feature. From our point of view, it is necessary that we should focus on the semantic loss of foreground object, not all components or a special case (e.g. faces).

III. FOREGROUND MEASUREMENT

Most existing studies adopt global saliency map to produce an overall quality assessment while they neglect the particularity of foreground object. Since the main factors of perceptual quality for images with and without foreground objects are quite different, the classification of images is necessary by judging whether they contain foreground objects. Therefore, we propose to conduct foreground object detection, and adopt two complementary features to measure the quality of retargeted images.

We first perform foreground detection on original images, and decide whether images contain obvious foreground object. If images have an obvious foreground object, we respectively extract the foreground object in original images and retargeted images, and design a foreground measurement as follows: 1) a special designed input adaption is used to produce the foreground object patches from both original images and retargeted images, and two foreground object patches are fed to a pre-trained CNN respectively to extract their semantic features. The similarity of these two semantic features is our semantic similarity measurement. 2) we calculate the ratio of foreground object pixels in

(a) M_{fod} of images with obvious foreground object. (b) M_{fod} of images without obvious foreground object.

FIGURE 2. Comparison of M_{fod} of images with and without obvious foreground object.

original images and retargeted images to represent our size change ratio feature.

A. FOREGROUND OBJECT DETECTION

Saliency maps adopted in previous IRQA are not directly applied into the foreground object detection because they only present coarse relative importance of different pixels. Therefore, these traditional saliency detection methods fail to generate an accurate foreground object detection result. By contrast, some attention mechanism based convolutional neural networks (CNN) have achieved great success on saliency detection. In our work, we introduce a saliency model, i.e. PiCANet [10], to detect the foreground object, which detects salient objects more accurately when comparing with other deep learning methods [34]–[36]. We apply PiCANet to extract the image saliency map in IRQA benchmark datasets, and some foreground object detection results are shown in Fig. 2. It is obvious that image saliency maps with obvious foreground object have larger saliency values in foreground areas than those without obvious foreground object. To distinguish images with and without foreground object, we calculate the mean saliency value of pixels in salient area, and select a threshold η to judge. The mean saliency value of salient area, denoted as S_{fod}, is calculated by:

$$S_{fod} = \frac{\sum_{\boldsymbol{p}\in M_{fod}, S(\boldsymbol{p})>0} S(\boldsymbol{p})}{N} \tag{1}$$

where the foreground object detection map M_{fod} is the resized output of PiCANet, $\boldsymbol{p}$ represent the pixel position and saliency value in position $\boldsymbol{p}$ is denoted as $S(\boldsymbol{p})$. We only consider the mean of pixel saliency value, and thus N is the total number of pixels with non-zero saliency values. Then we can judge whether the original image contains an obvious foreground object depending on whether $S_{fod} > \eta$. In our work, we set $\eta = 60$ empirically.

B. SEMANTICS SIMILARITY MEASUREMENT

In recent years, convolutional neural networks (CNN) has achieved great success in many high-level semantic related tasks [37]–[42]. CNN can absorb knowledge from quantities of data, and can learn to extract semantic features for image classification. The idea of encoding the semantic information using the hidden layer output of CNN is actually widely used in the area of face detection [43] and person re-identification [44]. Therefore, we propose to introduce CNN to extract the semantic information of foreground object for further analysis. However, it remains a challenging problem to directly apply CNN to solve IRQA. On the one hand, CNN generally requires input images to have a fixed size, but the change of aspect ratio itself is a crucial factor to affect the quality of retargeted images. Therefore, it is not acceptable to directly resize both original image and retargeted image to the same specific size. On the other hand, a large amount of data is required for training CNN, but the existing IRQA datasets contain few data due to the high cost of manually labeling. Besides, the overall quality of a retargeted image relies on not only itself but also the comparison with its corresponding original image.

To overcome this limitation, we first design a input adaption method to meet the same size requirement of network without changing the aspect ratio. After the input adaption, we use a pre-trained VGG16 network [37] to extract semantic features from original and retargeted images. Here VGG16 network is a pre-trained CNN on IMAGE-NET [45] which contains images of 1000 categories. In [46], experimental results indicate that higher layers of CNN represent the high-level semantic information. In our work, the

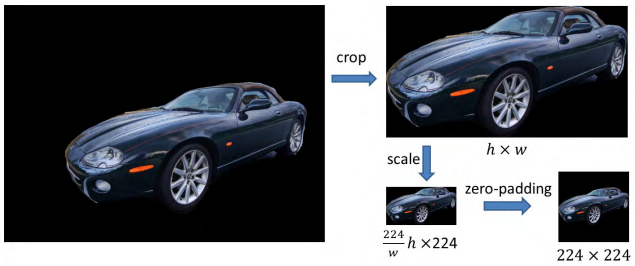


FIGURE 3. The preprocessing procedure for foreground object detection.

penultimate layer output of VGG16 network, also denoted as fc7, is adopted to represent the semantic feature vectors. Finally, the cosine similarity of semantic feature vectors between original and retargeted images is calculated to measure the perceptual quality of foreground object in retargeted images.

1) NETWORK INPUT ADAPTION

The foreground object mask for original image is generated by the image binarization of M_{fod}, where the binarized threshold is given empirically as $\delta = 100$. we use M^o_{fod} to denote the binarized result of M_{fod}. For comparison with retargeted images, the foreground object mask for retargeted images, denoted as M^r_{fod}, is generated from M^o_{fod} based on the backward registration. For each of them, we crop the foreground object with the minimum available rectangle. Finally, we scale the cropped foreground object images and conduct zero-padding to attain the 224 × 224 output images, which is suitable for VGG16 network input. This process is shown in Fig. 3.

2) SEMANTIC QUALITY METRIC

Following many effective works in face recognition and person re-identification, e.g. [43], [44], we adopt the penultimate layer output (denoted as the semantic feature vector) as a good representation of semantic information. These works measure the similarity of semantic feature vectors to judge whether two images contains the same person, which inspires us to measure the similarity of semantic information to judge the information loss in the same way. Therefore, for an original image and the corresponding retargeted image, we perform the input adaption, and feed them to the pre-trained VGG16 network respectively. The penultimate layer outputs of two networks are considered as the semantic feature vectors, and then we calculate the cosine similarity of these two vectors to measure the semantic quality of foreground object in retargeted images. The semantic quality Q_{sem} is denoted as:

$$Q_{sem} = \frac{\boldsymbol{f}^o \cdot \boldsymbol{f}^r}{\|\boldsymbol{f}^o\| \, \|\boldsymbol{f}^r\|} \tag{2}$$

where $\boldsymbol{f}^o$ and $\boldsymbol{f}^r$ is the semantic feature vector for original and retargeted images respectively, and $\|.\|$ is the l_2 norm.

Fig. 4 shows the predicted semantic quality Q_{sem} by our method for 'butterfly' image set and 'car' image set. As for

FIGURE 4. Comparison of semantic similarity Q_{sem}. (a) and (b) are the extracted foreground object of original images of set butterfly and car respectively. (c) ∼ (j) are the extracted foreground object of retargeted images of set butterfly. (k) ∼ (r) are the extracted foreground object of retargeted images of set car. The corresponding retargeting operator name for each image is given, followed by its Q_{sem} in parentheses.

set 'butterfly', we notice that the scores of SCL and WARP are obviously lower than others while the score for CR is 1. This indicates that squeezing and cropping of foreground object both lead to significant quality reduction. Meanwhile, the score for CR is 1 because the foreground object is not squeezed or cropped at all. Actually, distortion and discontinuity are also two factors that lead to quality reduction. As for set 'car', SC doesn't squeeze the car as much as SCL, while the score of the retargeted image is still lower than that using SCL because of distortion. Similarly, SM barely squeezes the car, but the score of the retargeted image is also lower than that of CR because of discontinuity. In summary, our semantic similarity measurement is related to squeezing, cropping, distortion and discontinuity at the same time, and the final score of the semantic similarity measurement can be considered as a comprehensive assessment for these four factors.

C. SIZE CHANGE RATIO MEASUREMENT

Although the semantic quality is an effective metric because it takes many factors into account such as squeezing, cropping,

distortion and discontinuity, it ignores the image scaling of foreground object. Some retargeting methods such as SV and SNS prevent the aspect ratio of foreground object from changing by shrinking it, which usually leads to quality reduction. Therefore, we adopt the ratio of foreground object pixels between original with retargeted images to represent the change of size by:

$$Q_{size} = \frac{\left| \boldsymbol{p} \subseteq M^r_{fod} \right|}{\left| \boldsymbol{p} \subseteq M^o_{fod} \right|} \tag{3}$$

where $|\cdot|$ stands for the number of the set.

As shown in Fig. 5, Q_{sem} of operator SV, SC and SNS are 0.9946, 0.9909 and 0.9964 respectively, and they are very close. However, these three retargeted images have different votes. The semantic quality values are not consistent with votes, but the size change ratio Q_{size} are highly consistent with votes. Therefore, we suggest that Q_{sem} and Q_{size} are two complementary features, and they are both necessary for measuring the foreground object quality in IRQA.

IV. GLOBAL MEASURES

Besides the quality of foreground object, and the global quality is also an important factor. Moreover, some images do not have obvious foreground object at all. Therefore, it is necessary to measure the global quality of retargeted images as well. In [12], ARS and EGS are two effective global features, and thus we adopt these two global features, and improve ARS to adapt our framework.

A. ASPECT RATIO SIMILARITY

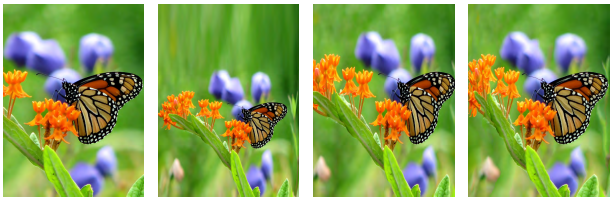
ARS is a simple but effective low-level feature which mainly considers structure loss and content loss of local blocks. First of all, the original image is partitioned into squares of pixel size 16×16. According to the backward registration results, the mapped pixel sets corresponding to each square are then attained. To measure the quality of each block in the original image, the quality of each block s_{ar} is given by:

$$s_{ar} = \left[\frac{2 \cdot r_w \cdot r_h + C}{r_w^2 + r_h^2 + C} \right] \cdot \left[e^{-\alpha(r_m - 1)^2} \right] \tag{4}$$

where r_w and r_h are width and height change ratios of bounding boxes, and $r_m = (r_w + r_h)/2$ is regarded as the absolute size change. C is a small constant for stability and we choose $C = 10^{-6}$. In Eq.4, the left part concentrates on aspect ratio change while the right part considers the affect of content loss. These two complementary measures are combined using a parameter α to adjust the balance between structure loss and content loss. A larger α results in more penalty for content loss. In [12], the quality of ARS Q_{ARS} is calculated by the weighted sum of the quality of each block s_{ar},

$$Q_{ARS} = \sum_{i=1}^{N} w(i) s_{ar}(i) \tag{5}$$

(a) The original image of set butterfly



(b) CR
Votes:52
Q_{sem}:1.0000
Q_{size}:0.9950

(c) SV
Votes:34
Q_{sem}:0.9946
Q_{size}:0.3661

(d) MOP
Votes:26
Q_{sem}:0.9897
Q_{size}:0.7193

(e) SC
Votes:44
Q_{sem}:0.9909
Q_{size}:0.8684

(f) SCL
Votes:6
Q_{sem}:0.9602
Q_{size}:0.5029

(g) SM
Votes:40
Q_{sem}:0.9905
Q_{size}:0.9475

(h) SNS
Votes:40
Q_{sem}:0.9964
Q_{size}:0.4292

(i) WARP
Votes:10
Q_{sem}:0.8394
Q_{size}:0.4013

FIGURE 5. Votes, Q_{sem} and Q_{size} of retargeted images in set 'butterfly'. (a) is the original image of set 'butterfly'. (b) ~ (i) are the retargeted images of different retargeting operator. The corresponding retargeting operator name for each image is given, followed by its votes, Q_{sem} and Q_{size}.

where $w(i)$ is the saliency weight for the i^{th} block, and it is evaluated by summing the normalized pixel-wise saliency values in the i^{th} block.

In [12], a saliency extraction method [48] is adopted. From our observation, it is not very suitable to match the requirement of ARS for images without obvious foreground object. As shown in Fig. 6, saliency values of saliency map [48] mainly concentrate on the local region, and it generally leads to neglecting the importance of global quality when evaluating images without obvious foreground object. On the contrary, the saliency values of GBVS [47] are smoother and more uniform in the global region. The more saliency values exist in the global region, the more effective quality score is generated for images without obvious foreground object. Therefore, we replace the saliency extraction method [48] with GBVS when evaluating images without obvious foreground object.

B. EDGE GROUP SIMILARITY

Broken and distorted lines and edges are two common reasons why people dislike a retargeted image [26]. Distortion of straight lines and smooth curves are usually the decisive factors for the overall quality. Sometimes the contour distortion also leads to artifacts that cause serious deterioration to

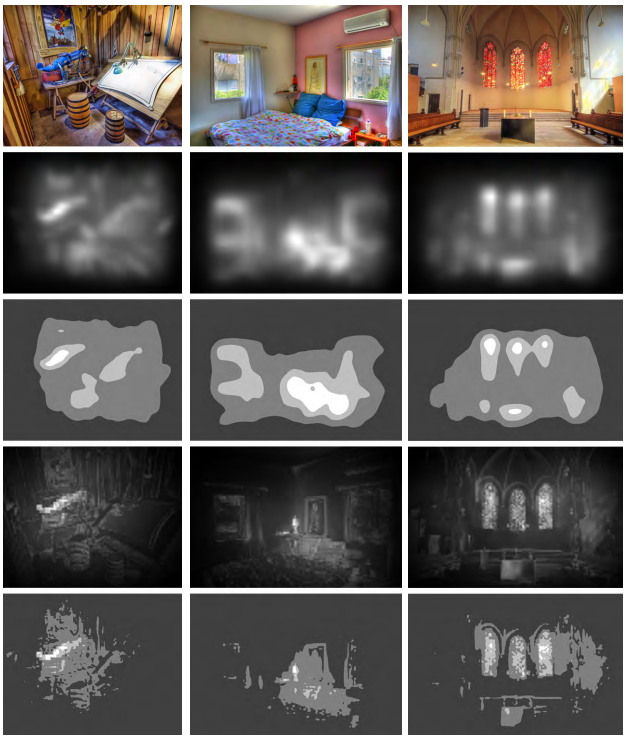


FIGURE 6. Comparison of GBVS [47] and the saliency extraction method [48]. The first row contains 3 original images without obvious foreground object. The second row is the GBVS saliency maps and the third row is the quantized saliency maps of GBVS saliency maps. The fourth row is the saliency maps proposed in [48] and the fifth row is the quantized saliency maps of [48]. The quantization level for both saliency maps are (0,0.25], (0.25,0.5], (0.5,0.75], (0.75,1] and a darker region correspond to a lower saliency range.

perceptual quality. Therefore, the edge distortion should be modeled into IRQA.

The following part is a brief introduction of the implementation to measure edge group similarity [12]. The initial edge map is first generated using the structure edge detector in [49]. The non-maximum suppression (NMS) [50] and edge group clustering procedures [51] are performed to obtain sparse edge group representation. For each edge group EG'_k in the retargeted image, the corresponding edge group EG_k in the original image is matched according to the backward registration results. Chamfer matching [52] is then applied to calculate the distance between edge group maps of original image and retargeted image. The chamfer distance is denoted as $d_{CM}\left(EG_k, EG'_k\right)$. Finally, the overall edge group similarity is calculated by

$$Q_{EGS} = e^{-\beta\sqrt{\frac{1}{N}\sum_{k=1}^{N} d_{CM}\left(EG_k, EG'_k\right)}} \tag{6}$$

where β can adjust the distribution of score and N is the number of edge groups.

V. EXPERIMENT

We conduct experiments on two databases, i.e. RetargetMe and CUHK, and experimental results show that our proposed method provides state-of-the-art performance. Introduction and performance on both databases are given as follows. In addition, feature and parameter study are given for further analysis.

A. DATASET INTRODUCTION

1) RETARGETME

MIT RetargetMe database contains 37 sets of images. Each set contains an original image and eight corresponding retargeted images using different operators. Retargeting operators including CR, SV [2], MOP [5], SC [1], SCL, SM [6], SNS [3], WARP [4] are applied to generate retargeted images with 50% or 25% size reduction in height or width. Votes for each retargeted image are collected in the way of pair-wise comparison. Volunteers are given two retargeted images in the same image set, and required to vote for retargeted images of better quality. In a certain image set, any two operators have been compared, and each pair of images appears the same frequency. Therefore, the collected total votes for each retargeted image can represent its subjective perceptual quality.

We follow previous works and adopt Kendall rank correlation coefficient (KRCC) [53] to measure the correlation between subjective rank and objective rank.

$$KRCC = \frac{N_r - N_d}{\binom{N}{2}} \tag{7}$$

where N_r and N_d are consistent pairs number and inconsistent pairs number respectively, and $N = N_r + N_d$. If the objective rank is completely consistent with subjective rank, then N_r equals the total number of compared pairs and thus $KRCC = 1$. If two ranks are completely inconsistent, on the contrary, we have $KRCC = -1$. We should notice that these two cases are essentially the same because we can arbitrarily choose whether a higher or a lower predicted score represents a higher perceptual quality. When a proposed method gives a totally uncorrelated quality prediction with the subjective rank, theoretically we will have $N_r = N_d$ and thus $KRCC = 0$.

2) CUHK

CUHK database [55] contains 57 sets of images. Each set contains an original image and three corresponding retargeted images. Besides the eight operators in RetargetMe database, optimized seam-carving and scale [7] and energy-based deformation [8] are also applied. Different from RetargetMe database, three retargeted images in the same set possibly have different reduction in size. Meanwhile, five-level scores are collected to generate Mean Objective Scores (MOS) instead of the preference of image pairs.

For CUHK database, subjective scores and objective scores are compared using four metrics: Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), root mean squared error (RMSE) and outlier ratio (OR). The outlier ratio represents the percentage of retargeted images whose mapped scores are not inside the interval $[MOS - 2\sigma, MOS + 2\sigma]$, where MOS and σ are

TABLE 1. Performance comparison on RetargetMe.

Method	Attribute						Total	
	Lines edges	Faces people	foreground objects	Texture	Geometric structures	Symmetry	mKRCC	stdKRCC
BDS [24]	0.04	0.19	0.167	0.06	-0.004	-0.012	0.083	0.268
EH [22]	0.043	-0.076	-0.079	-0.060	0.103	0.298	0.004	0.334
SIFT flow [9]	0.097	0.252	0.218	0.161	0.085	0.071	0.145	0.262
EMD [25]	0.22	0.262	0.226	0.205	0.237	0.5	0.251	0.272
PGDIL [54]	0.431	0.39	0.389	0.286	0.438	0.523	0.415	0.296
ARS [11]	0.463	0.519	0.444	0.33	0.505	0.464	0.452	0.283
MLF [12]	0.486	**0.605**	0.544	0.384	0.536	0.536	0.512	0.251
HCnDL [13]	0.497	0.472	0.468	0.393	0.545	**0.631**	0.494	0.243
OUR	**0.521**	0.598	**0.619**	**0.473**	**0.554**	0.548	**0.552**	**0.221**

TABLE 2. Performance comparison of different sets on RetargetMe.

Method	Set		Attribute						Total	
	D_f	D_{nf}	Lines edges	Faces people	foreground objects	Texture	Geometric structures	Symmetry	mKRCC	stdKRCC
MLF [12]	✓		**0.536**	0.566	0.544	0.417	0.536	**0.607**	0.544	0.241
		✓	0.429	0.429	-	0.179	0.5	0.5	0.396	**0.238**
	✓	✓	0.486	**0.605**	0.544	0.384	0.536	0.536	0.512	0.251
OUR	✓		0.505	**0.607**	**0.619**	**0.5**	**0.557**	0.536	**0.577**	**0.198**
		✓	**0.543**	**0.536**	-	**0.393**	**0.548**	**0.554**	**0.494**	0.268
	✓	✓	**0.521**	0.598	**0.619**	**0.473**	**0.554**	**0.548**	**0.552**	**0.221**

the mean and the deviation of objective scores respectively. An effective objective quality assessment method generally has high PLCC and SRCC while RMSE and OR are contrarily low. In [55], there is a non-linear mapping between subjective scores and objective scores. The mapping function is shown as:

$$f(x) = \beta_1\left(\frac{1}{2} - \frac{1}{e^{\beta_2(x-\beta_3)}}\right) + \beta_4 x + \beta_5 \tag{8}$$

where $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are mapping parameters produced by non-linear fitting and the mapping target is fitting subjective scores with objective scores.

B. PARAMETER SETTINGS

We empirically set the threshold η ($\eta = 60$) for estimating whether an image has obvious foreground object and the binarization threshold δ ($\delta = 100$) for generating the mask M^o_{fod} from the foreground object M_{fod}. Since we switch from the original saliency map [48] to GBVS [47], we provide a study on the effect of different α in RetargetMe. As shown in Fig. 7, ARS based on GBVS is generally more effective than based on the original saliency map [48] when α changes from 0 to 1 with the stride of 0.1. For saliency map [48], the highest mean KRCC of D_{nf} is 0.435 when $\alpha = 0.3$, while for GBVS, the highest mean KRCC is 0.494 when $\alpha = 0.9$. As for EGS, we follow [12] and set $\beta = 0.2$.

C. PERFORMANCE ON RETARGETME

We follow MLF [12] and adopt SVM^{rank} [56] in the way of leave one-out cross-validation (LOOCV) to attain the objective rank. Based on our foreground object detection result, the original 37 images are divided into two groups, i.e. images D_f with obvious foreground object (26 images) and images D_{nf} without obvious foreground object (11 images). Radial Basis Function (RBF) kernel is applied in SVM^{rank} and we follow the parameters in open-source code of [12] where $\gamma = 2^{3.2}$ and $C = 2^{4.8}$.

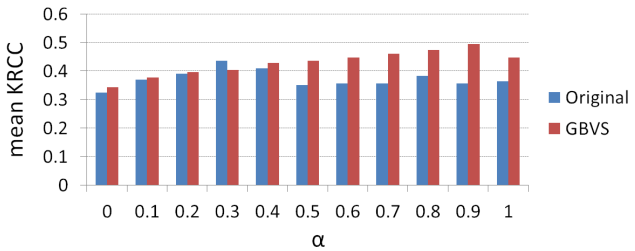


FIGURE 7. Mean KRCC on D_{nf} of retargetMe using different saliency map and different α. The original saliency map is from [48] and GBVS is from [47].

The mean and standard deviation of KRCC are presented in TABLE 1. As shown in TABLE 1, when comparing our method with existing best methods, mean KRCC values on attribute *"Foreground Objects"* and *"Texture"* have been improved significantly by 13.8% and 20.4% respectively. We believe that our foreground measurement contributes to the improvement on attribute *"Foreground Objects"*, and our adjustment of image saliency maps contributes to the improvement on attribute *"Texture Objects"*. When comparing the overall performance, the overall mean KRCC of our method is 0.552, which is 7.8% higher than that of state-of-the-art method (MLF [12]), whose mean KRCC is 0.512.

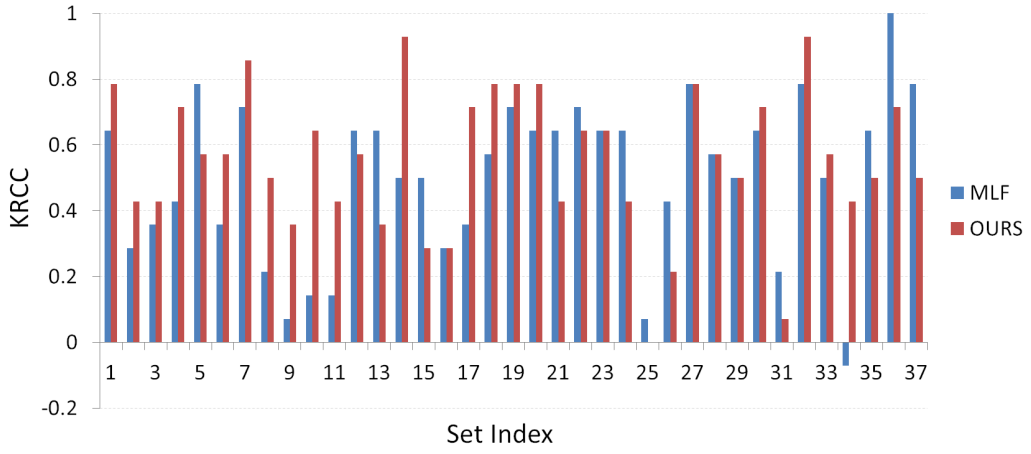


FIGURE 8. KRCC of each image set in RetargetMe.

TABLE 3. Performance Comparison of two cases on RetargetMe. Case I: Top 3 subjective ranking and top 3 objective ranking. Case II: Top 3 subjective ranking.

Method	case I		case II	
	mean KRCC	std KRCC	mean KRCC	std KRCC
BDS [24]	0.108	0.532	0.132	0.372
EH [22]	-0.071	0.593	0.009	0.413
SIFT flow [9]	0.298	0.483	0.255	0.348
EMD [25]	0.326	0.496	0.339	0.442
PGDIL [54]	0.533	0.383	0.532	0.28
ARS [11]	0.45	0.469	0.52	0.347
MLF [12]	0.518	0.459	0.589	0.310
OUR	**0.634**	**0.381**	**0.664**	**0.265**

To further analyze the performance of different images set, i.e. D_f and D_{nf}, the subset performance comparison between [12] and our method are given in TABLE 2 and Fig.8 as well. As shown in TABLE 2 and Fig.8, we achieve great improvement on both sets compared to MLF. We check the foreground object detection result manually, and find that all 18 sets of attribute *"Foreground Objects"* are included in D_f. Therefore, we are convinced that our foreground measures contribute to the great improvement on the attribute of *"Foreground Objects"*, and it proves the effectiveness of our specifically designed measures for foreground object.

We also follow the two types of top3 ranking KRCC evaluation in [12]. Case I considers only the pairs $s.t.$ $(rank_s(i) \leq 3 \vee rank_s(j) \leq 3) \wedge (rank_o(i) \leq 3 \vee rank_o(j) \leq 3)$ and case II considers those $s.t.$ $rank_s(i) \leq 3 \vee rank_s(j) \leq 3$. Here i, j are indices of the two retargeted images to compare, and $rank_s$, $rank_o$ represent the subjective rank and objective rank respectively. Both evaluations skip the comparison within retargeted images whose rank is low. As shown in TABLE 3, the overall mean KRCC of our method is 0.634 in the case I, which is 18.9% higher than the best traditional method PGDIL [54], and the overall mean KRCC of our method is 0.664 in the case II, which is 12.7% higher than the best traditional method MLF [12].

TABLE 4. Performance on CUHK.

Method	D_f	D_{nf}	PLCC	SRCC	RMSE	OR
BDS [24]	-	-	0.2896	0.2887	12.922	0.2164
EH [22]	-	-	0.3422	0.3288	12.686	0.2047
SIFT flow [9]	-	-	0.3141	0.2899	12.817	0.1462
EMD [25]	-	-	0.276	0.2904	12.977	0.1696
PGDIL [54]	-	-	0.5403	0.5409	11.361	0.152
ARS [11]	-	-	0.6835	0.6693	9.855	0.0702
HCnDL [13]	-	-	0.7170	0.6847	9.1352	0.0215
MLF [12] (train on D_f and D_{nf})	✓		0.7602	0.7511	9.2016	**0.0732**
		✓	0.7411	**0.7321**	7.8178	0
	✓	✓	0.7577	0.7383	**8.525**	**0.0294**
MLF [12] (train on D_f or D_{nf})	✓		0.7487	0.7414	9.385	**0.0732**
		✓	0.7542	0.7206	7.6313	0
OUR	✓		**0.7759**	**0.7516**	**8.9318**	0.0894
		✓	**0.7732**	0.7169	**7.3719**	0
	✓	✓	**0.7745**	**0.7706**	8.5421	0.0702

Another metric proposed in [12] is rank-n accuracy, which aims to examine the ability to select the most favored retargeted image. Here, rank-n accuracy is the percentage of sets where the top-n ranked retargeted images include the retargeted image with most subjective votes. As shown in Fig. 9, the most favored retargeted images in 54.1% sets have the highest objective scores using our method, and the improvement is huge when comparing with ARS [11], whose rank-1 accuracy is only 29.7%. Although the rank-2 accuracy of MLF [12] is 56.7%, we achieve an even better rank-2 accuracy, which is 81.1%. The rank-3 accuracy of our method shows that more than 90% sets are included in top-3 objective scored retargeted images, which means that our method has a strong ability to pick out the most subjectively liked retargeted images.

D. PERFORMANCE ON CUHK

When testing on CUHK, we are more concerned about fitting the subjective scores with the objective scores. Therefore, we follow the five-fold cross-validation training using SVR applied in [12] to generate the fused objective scores.

TABLE 5. Feature analysis on D_f of RetargetMe.

Feature				Attribute						Total	
Q_{ars}	Q_{egs}	Q_{sem}	Q_{size}	Lines edges	Faces people	foreground objects	Texture	Geometric structures	Symmetry	mKRCC	stdKRCC
✓				0.378	0.398	0.401	0.321	0.421	0.536	0.407	0.255
	✓			0.311	0.306	0.27	0.3	0.343	0.393	0.297	0.216
✓	✓			0.5	0.566	0.508	0.452	0.514	**0.571**	0.519	0.216
✓	✓	✓		0.505	0.571	0.564	0.405	0.557	**0.571**	0.552	0.24
✓	✓		✓	**0.515**	0.571	0.599	**0.512**	**0.607**	**0.571**	0.57	0.221
		✓		0.245	0.316	0.393	0.179	0.379	0.357	0.346	0.357
			✓	0.235	0.372	0.389	0.345	0.329	0.321	0.338	0.267
		✓	✓	0.281	0.434	0.472	0.31	0.371	0.429	0.404	0.281
✓		✓	✓	0.495	0.561	0.603	0.441	0.571	0.536	0.558	0.237
	✓	✓	✓	0.337	0.434	0.460	0.345	0.414	0.464	0.426	0.213
✓	✓	✓	✓	0.505	**0.607**	**0.619**	0.5	0.557	0.536	**0.577**	**0.198**

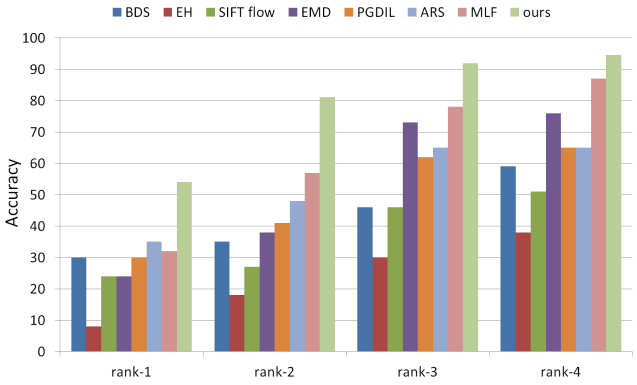


FIGURE 9. Top-N rank accuracy.

TABLE 6. Feature analysis on D_f of CUHK.

Q_{ars}	Q_{egs}	Q_{sem}	Q_{size}	PLCC	SRCC	RMSE	OR
✓				0.6772	0.6595	10.4172	0.1057
	✓			0.4398	0.4547	12.7158	0.1707
✓	✓			0.6969	0.6915	10.155	0.122
✓	✓	✓		0.7374	0.7295	9.5643	0.1138
✓	✓		✓	0.7185	0.7036	9.8489	0.0976
		✓		0.6061	0.6563	11.2609	0.1382
			✓	0.7164	0.6909	9.8775	0.1057
		✓	✓	0.7472	0.7186	9.4149	0.0976
✓		✓	✓	0.7741	0.7483	8.9632	0.0976
	✓	✓	✓	0.7459	0.7193	9.432	0.0976
✓	✓	✓	✓	**0.7759**	**0.7516**	**8.9318**	**0.0894**

The training procedure is conducted on D_f and D_{nf} respectively using different combination of features. Since features adopted in D_f and D_{nf} are different, we don't use the same parameter setting for these two sets. The optimal parameters γ and C are chosen by grid search for these two sets. In D_f, $\gamma = 2^{-1.2}$ and $C = 2^{0.8}$. In D_{nf}, $\gamma = 2^{2.7}$ and $C = 2^{3.9}$. To reproduce the experiment results of [12], we adopt the parameter setting $\gamma = 2^{3.1}$ and $C = 2^{4.3}$. According to our foreground object detection results, D_f consists of 41 images and D_{nf} consists of 16 images in CUHK. Following the experimental setting in [12], we conduct five-fold cross-validation training on D_f and D_{nf} individually, and then take the median performance of 1000 times train-test random split. In order to compare the performance of [12] with our method, we have attempted to reproduce the experiment result in [12]. The results are produced by training 1) on the whole CUHK dataset, i.e. on both D_f and D_{nf}, 2) on D_f individually, 3) on D_{nf} individually. In TABLE 4 we present the performance of other methods. The result shows that our proposed method outperforms all other methods, and it has great improvement on D_f and D_{nf} in PLCC compared to [12]. Although we have a higher OR compared to [12], the outlier coefficient in [55] shows that the subjective assessments of different people are inherently not totally consistent. In CUHK database, 15 images are recognized as the outlier following the subjective assessments of different people, i.e. $OR = 0.0877$. In our experiment, 12 images are considered as the outlier, i.e. $OR = 0.0702$. Therefore, our performance of OR is acceptable.

E. FEATURE ANALYSIS

To further study the effect of each single feature and the complementarity of features, we test on each single feature and some designed combination of features. Results of RetargetMe database and CUHK database are shown in TABLE 5 and TABLE 6. As shown in TABLE 5, the best performance for attributes lines/edges, texture, geometric structures and symmetry is achieved using the combination of Q_{ars}, Q_{egs} and Q_{size} while the best performance for the whole set, attributes faces/people and foreground objects is achieved using all four features. This indicates that Q_{sem} and Q_{size} have their own advantages on image sets of different attributes. In [13], the semantic measurement is also proposed, the mean KRCC of semantic measurement is only -0.0676. However, the mean KRCC of our semantic measurement Q_{sem} is 0.346. It indicates that our proposed semantic measurement is much more effective than semantics similarity measurement proposed in [13]. If only using the feature setting Q_{ars} and Q_{egs}, the mean KRCC of RetargetMe is 0.519 on set D_f. The improvement ratio is 6.4% after adding feature Q_{sem}, and the improvement ratio is 11.1% after adding both Q_{sem} and

Q_{size}. Similarly, as shown in TABLE 6, PLCC of CUHK is 0.6969 when using the feature setting Q_{ars} and Q_{egs}. The improvement ratio is 5.8% after adding feature Q_{sem}, and the improvement ratio is 11.3% after adding both Q_{sem} and Q_{size}. Therefore, we conclude that the foreground and global measurement are indeed complementary to solve IQRA.

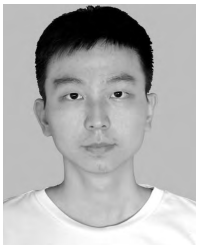
VI. CONCLUSION

This paper proposed a content-aware image retargeting assessment method using foreground and global measurement to provide objective scores for retargeted images. First, the backward registration [9] was used to estimate the pixel-level correspondence between the original and retargeted images. Then, images were divided into two categories according to the foreground object detection result, and different corresponding measurements were designed for them. For those with obvious foreground object, both foreground and global measurement were applied. For others, only global measurement was conducted. Foreground measurement included two complementary features of different levels: the high-level semantic similarity feature and low-level size ratio feature. When extracting the high-level semantic similarity features, a network input adaption method was specially designed to avoid the semantic information loss, and two pre-trained convolutional neural networks were adopted because of small scale data in IRQA. The low-level feature was the size change ratio of foreground object between original images and retargeted images. Global features included the improved ARS and EGS, whom were weighted by the saliency map of GBVS. Finally, a learned model was used to predict the perceptual quality of retargeted images. When experimenting on two public databases (i.e. RetargetMe and CUHK), our method achieved state-of-the-art performance comparing with other existing methods. For further improvement, we thought that we should pay more attention to semantic content analysis to achieve better performance in IRQA.

REFERENCES

[1] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 16:1–16:9, Aug. 2008.
[2] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. no. 126.
[3] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, p. 118, Dec. 2008.
[4] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–6.
[5] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, p. 23, 2009.
[6] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2009, pp. 151–158.
[7] W. Dong, N. Zhou, J.-C. Paul, and X. Zhang, "Optimized image resizing using seam carving and scaling," *ACM Trans. Graph.*, vol. 28, no. 5, p. 125, Dec. 2009.
[8] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1257–1268, 2009.
[9] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.
[10] N. Liu, J. Han, and M.-H. Yang, "Picanet: Learning pixel-wise contextual attention for saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 3089–3098.
[11] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration-based aspect ratio similarity for image retargeting quality assessment," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4286–4297, Sep. 2016.
[12] Y. Zhang, W. Lin, Q. Li, W. Cheng, and X. Zhang, "Multiple-level feature-based measure for retargeted image quality," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 451–463, Jan. 2018.
[13] Z. Fu, F. Shao, Q. Jiang, R. Fu, and Y.-S. Ho, "Quality assessment of retargeted images using hand-crafted and deep-learned features," *IEEE Access*, vol. 6, pp. 12008–12018, 2018.
[14] Y. Kao, R. He, and K. Huang, "Deep aesthetic quality assessment with semantic information," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1482–1495, Mar. 2017.
[15] M. Oszust, "Optimized filtering with binary descriptor for Blind image quality assessment," *IEEE Access*, vol. 6, pp. 42917–42929, 2018.
[16] Y. Ding, R. Deng, X. Xie, X. Xu, Y. Zhao, X. Chen, and A. S. Krylov, "No-reference stereoscopic image quality assessment using convolutional neural network for adaptive feature extraction," *IEEE Access*, vol. 6, pp. 37595–37603, 2018.
[17] L. Guo, C. Guo, L. Lei, Q. Huang, Y. Li, and X. Li, "Two-stage local constrained sparse coding for fine-grained visual categorization," *Sci. China Inf. Sci.*, vol. 61, no. 1, 2018, Art. no. 018104.
[18] L. Li, Y. Yan, Z. Lu, J. Wu, K. Gu, and S. Wang, "No-reference quality assessment of deblurred images based on natural scene statistics," *IEEE Access*, vol. 5, pp. 2163–2171, 2017.
[19] J. Ma, P. An, L. Shen, and K. Li, "Reduced-reference stereoscopic image quality assessment using natural scene statistics and structural degradation," *IEEE Access*, vol. 6, pp. 2768–2780, 2018.
[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[21] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
[22] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[23] E. Kasutani and A. Yamada, "The MPEG-7 color layout descriptor: A compact image feature description for high-speed image/video segment retrieval," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 1, Oct. 2001, pp. 674–677.
[24] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
[25] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. ICCV*, vol. 9, Sep./Oct. 2009, pp. 460–467.
[26] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graph.*, vol. 29, no. 5, pp. 160:1–160:10, 2010.
[27] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, and C.-W. Lin, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 4, no. 1, pp. 95–105, Mar. 2014.
[28] Y. Liang, Y.-J. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 2, pp. 1099–1110, Feb. 2016.
[29] Y. Chen, Y.-J. Liu, and Y.-K. Lai, "Learning to rank retargeted images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3994–4002.
[30] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning sparse representation for objective image retargeting quality assessment," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1276–1289, Apr. 2018.
[31] S. A. Oliveira, S. S. Alves, J. P. Gomes, and A. R. R. Neto, "A bi-directional evaluation-based approach for image retargeting quality assessment," *Comput. Vis. Image Understand.*, vol. 168, pp. 172–181, Mar. 2018.
[32] Y. Zhang, K. N. Ngan, L. Ma, and H. Li, "Objective quality assessment of image retargeting by incorporating fidelity measures and inconsistency detection," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5980–5993, Dec. 2017.
[33] Z. Chen, J. Lin, N. Liao, and C. W. Chen, "Full reference quality assessment for image retargeting based on natural scene statistics modeling and bi-directional saliency similarity," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5138–5148, Nov. 2017.

[34] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. Torr, "Deeply supervised salient object detection with short connections," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3203–3212.
[35] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li, and P.-M. Jodoin, "Non-local deep features for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6609–6617.
[36] T. Wang, A. Borji, L. Zhang, P. Zhang, and H. Lu, "A stagewise refinement model for detecting salient objects in images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4019–4028.
[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[38] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[39] L. Guo, "Self-paced multi-task joint sparse representation method," *IEICE Trans. Inf. Syst.*, vol. E101.D, no. 8, pp. 2115–2122, 2018.
[40] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[41] R. Girshick, "Fast R-cnn," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[42] L. Guo, "Self-paced learning with statistics uncertainty prior," *IEICE Trans. Inf. Syst.*, vol. 101, no. 3, pp. 812–816, 2018.
[43] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1701–1708.
[44] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proc. ACM Multimedia Conf. Multimedia Conf.*, Oct. 2018, pp. 274–282.
[45] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.
[46] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2013, pp. 818–833.
[47] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 545–552.
[48] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.
[49] P. Dollár and L. C. Zitnick, "Structured forests for fast edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1841–1848.
[50] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 184–203, Nov. 1987.
[51] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 391–405.
[52] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Aug. 1977, pp. 659–663.
[53] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, pp. 81–93, Jun. 1938.
[54] 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. Topics Signal Process.*, vol. 8, no. 3, pp. 377–389, Jun. 2014.
[55] 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. Topics Signal Process.*, vol. 6, no. 6, pp. 626–639, Oct. 2012.
[56] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.



YUWEI LI received the B.S. degree in electronics and information engineering from the South China University of Technology, where he is currently pursuing the master's degree in electronic and communication engineering. His research interests include machine learning and computer vision.

LIHUA GUO received the B.S. and M.S. degrees from the Nanjing University of Posts and Telecommunications (NUPT), in 1999 and 2002, respectively, and the Ph.D. degrees from Shanghai Jiao Tong University, in 2005. He is currently an Associate Professor with the South China University of Technology. His research interests are image understand and pattern recognition.

LIANWEN JIN was born in 1968. He received the B.Eng. degree from the Department of Electronics, University of Science and Technology of China (USTC), and the Ph.D. degree in communication and information system from the South China University of Technology (SCUT), in 1991 and 1996, respectively. He visited Motorola Research Center, in 2000, and The University of Hong Kong, in 2002, as a Research Fellow. He is currently a Professor with the School of Electronic and Information Engineering, South China University. He has published more than 70 papers in the field of handwritten character recognition, pattern recognition, neural networks, image processing, and intelligent systems. Since 1998, he has been the Principle Investigator of more than 15 research projects. He is a member of the IEEE Computer Society, the IEEE Communication Society, the China Image and Graphics Society, the China Communication Society (Senior), the Guangdong Image, and Graphics Council. He received the award of the New Century Excellent Talent Program of MOE, in 2006.