A Deep Evaluator for Image Retargeting Quality by Geometrical and Contextual Interaction

Bin Jiang, Jiachen Yang, *Member, IEEE,* Qinggang Meng, *Senior Member, IEEE,* Baihua Li, Wen Lu, *Member, IEEE*

***Abstract*—Image is compressed or stretched during the multi device displaying, which will have a very big impact on the perception quality. In order to solve this problem, a variety of image retargeting methods have been proposed for the retargeting process. However, how to evaluate the results of different image retargeting is a very critical issue. In various application system, the subjective evaluation method can not be applied on a large scale. So we put this problem in the accurate objective quality evaluation. Currently, most of the image retargeting quality assessment(IRQA) algorithms use simple regression methods as the last step to get the evaluation result, which are not corresponding with the perception simulation in human vision system. In this paper, a deep quality evaluator for image retargeting based on segmented stacked AutoEnCoder is proposed. Through the help of regularization, the designed deep learning framework can solve the overfitting problem. The main contributions in this framework are to simulate the perception of retargeted images in human vision system. Specially, it trains two separated stacked AutoEnCoder models based on geometrical shape and content matching. Then the weighting schemes can be used to combine the obtained scores from two models. Experimental results in three well known databases show that our method can achieve better performance than traditional methods in evaluating different image retargeting results.**

***Index Terms*—Image retargeting quality assessment, segmented stacked AutoEnCoder, perception simulation, geometrical shape, content match.**

I. Introduction

With the popularity of mobile Internet and mobile devices, images and videos need to be played in different resolutions. In this way, devices in special resolution need to resize the images to meet the requirements [1]. However, the quality of image retargeting results affect the quality of experience [2]. Currently, many image retargeting method have been presented, all of which are not universal for application situation [3]–[9]. Therefore, it is necessary to design a universal retargeting method. In this process, how to evaluate the results of image retargeting becomes a critical issue, which will be the main purpose for this paper.

This work was partially supported by National Natural Science Foundation of China (No. 61471260) and Natural Science Foundation of Tianjin (No. 16JCYBJC16000)*(Corresponding author: Jiachen Yang.)*

B. Jiang and J. Yang are with School of Electrical and Information Engineering, Tianjin University, Tianjin, China (e-mail: jiangbin@tju.edu.cn; yangjiachen@tju.edu.cn).

Q. Meng and B. Li are with the Department of Computer Science, Loughborough University, Loughborough, UK (email:q.meng@lboro.ac.uk; b.li@lboro.ac.uk).

W. Lu is with School of Electronic Engineering, Xidian University, Xi'an, China (e-mail: luwen@mail.xidian.edu.cn).

At present, there have been many research achievements in image quality evaluation (IQA). In [10], Wang *et al.* used structural similarity to measure the error visibility for image quality assessment. Sheikh *et al.* made use of the relationship between image information and visual quality [11]. A novel feature similarity (FSIM) method designed for image assessment was proposed by Zhang *et al.* [12]. Later, the image quality evaluation was gradually developed to video evaluation [13]. Based on these typical algorithms, we can arrive at a primary conclusion. In addition, some video quality assessment metrics were proposed [14], [15]. Traditional full reference image quality assessment often makes use of the subtraction between distortion image and reference image [16]. In the early researches in image retargeting evaluation, many researchers directly applied the traditional IQA algorithms on IRQA. However, there is a big difference between them. The most obvious problem is that image retargeting pays attention on the geometrical shape and contextual matching after resolution changes, which is difficult to solve by traditional IQA algorithms.

Based on these issues, many special evaluation algorithms for image retargeting were designed. Simakov *et al.* [17] proposed a principled approach based on optimization of well-defined similarity measure. The problem it considered is retargeting of image/video data into smaller sizes. In [18], Liu *et al.* presented an objective metric simulating the human vision system (HVS). Different from traditional objective assessment methods that work in bottom-up manner, it used a reverse order (top-down manner) that organizes image features from global to local viewpoints. Inspired by [10], Fang *et al.* proposed an effective but simple image retargeting quality assessment method which can be called as IR-SSIM [19]. In the assessment process, the SIFT-flow is used to find the pixel correspondence. And an SSIM map is computed to measure the preserved structure information in the retargeted image. Zhang *et al.* made another breakthrough by developing an aspect ratio similarity metric. In the computing process, local block quality changes are used [20]. Rubinstein *et al.* first set up a database for evaluating retargeting image named RerargetMe [21]. They present the first comprehensive perceptual study and analysis on image retargeting. Ma *et al.* put forward another database designed as a diverse independent public database with corresponding subjective scores, and they also give a effective evaluation method [22]. Based on perceptual geometric distortion and information loss, Hsu *et al.* presented a new objective quality assessment method for image retargeting. At the same time, they also built another

Fig. 1: Examples of retargeting operators. (a) Original image; (b) Simple Scaling Operator (SSO); (c) Manual Cropping (MC); (d)Seam Carving (SC); (e) Nonhomogeneous Warping (WARP); (f) Scale and Stretch (SCST); (g) Multi Operator (MULTI); (h) Shift Map (SM); (i) Streaming video (STVI).

database called NRID [23]. Jiang *et al.* made a research on IRQA through learning sparse representation. In [24], they focused on finding the potentiality of sparse presentation based on distortion sensitive features. Ma *et al.* resort to the pairwise rank learning approach to discriminate the perceptual quality between the retargeted image pairs [25]. Liang *et al.* considered five different key factors for image retargeting, such as salient regions, influence of artifacts, the global structure of the image, well-established aesthetics rules and preservation of symmetry [26]. Liu *et al.* put forward image retargeting quality assessment based on four quality factors and support vector regression [27]. They accounted quality factors into two categorizes: shape distortions and visual content changes.

Although the above algorithms have achieved some good results, they still have a lot of problems. These problems can be summarized in three aspects: 1) In IRQA, most of the evaluation methods use simple regression methods, which are not corresponding with the perception simulation in human vision system. Based on this consideration, the deep learning method will greatly benefit the accuracy of the evaluation algorithm. 2) Although there are some methods considering geometrical shape or contextual matching separately. The relationship between the features based on the two different parts has not been fully studied. If this problem can be solved, the evaluation result will be improved. 3) Most of the IRQA method only inherit traditional image evaluation framework. And how to design a framework designed specially for IRQA is very important. The cross database experiments are ignored in previous researches.

In the proposed algorithm, the following contributions of this paper can be summarized to improve the performance of IRQA algorithm.

1) Segmented stacked AutoEnCoder based on image representations is used to simulate the retargeting image perception process in human vision system. Specially, we propose a deep quality evaluator for image retargeting based on the connections of two modules: image representations and stacked AutoEnCoder. On the one hand, the image representations are used for the image information exaction, which can simulate the first image perception step from eyes to brains (V1 and V2 areas in visual pathway). On the other hand, stacked AutoEnCoder makes use of greedy training method layer by layer to train each layer of the network sequentially. The process is corresponding with the retargeting image perception in human brains (V4 area in visual pathway). Based on above consideration, we choose it to finish the final assessment of the retargeting images.

2) The proposed method overcomes overfitting and finds the complementary image features for whole framework. At present, the method of deep learning has promoted the ability of pattern recognition algorithm, but encounters overfitting problem in the traditional image quality evaluation. In order to solve the overfitting problem, we introduce regularization as the solution to make the deep model more accurate in the test stage. In addition, we choose the network input by considering two complementary parts: geometrical shape and contextual matching. And we make a deep study in the relationship between the two categories. Experiments show that the geometrical shape and content matching can actually provide more reliable feature group for deep learning on IRQA.

3) Cross database experiments have been done in this

research and it can promote the development in practical applications. In previous researches on IRQA, cross database experiments are always ignored during to the different building principles. In order to improve the practical value of image retargeting evaluation, we put forward the cross database experiments. On RetargetedMe and NRID, we set virtual DMOS for quality of every retargeted image based on the preferred number in original paired comparison methodology, which can be corresponding with CUHK. Through a comprehensive validation, the proposed metric correlates well with subjective observations and it can be used for general quality evaluator in image retargeting quality assessment.

The rest of this paper is organized as follows. Section II will illustrate the background and motivation based on the related work. In Section III, the special algorithm framework will be given. In Section IV and V, the experimental design and experimental results are shown. At last, conclusion will be given in Section VI.

II. Background and Motivation

A. From Traditional IQA to IRQA

The ultimate goal of image quality assessment is to simulate human perception and cognition process. In the specific evaluation environment, this simulation process will produce small differences. For traditional image quality assessment, the method design maximizes the relationship between quality of images and human visual system (HVS), which has yielded a lot of research results. In ventral stream of human visual system, the information passes through V1, V2 and V4. In this processing, it mainly considers shape recognition and object representation. In dorsal stream of human visual system, the information passes through V1, V2 and V5. In this processing, it mainly considers motion computation, object location and trajectory.

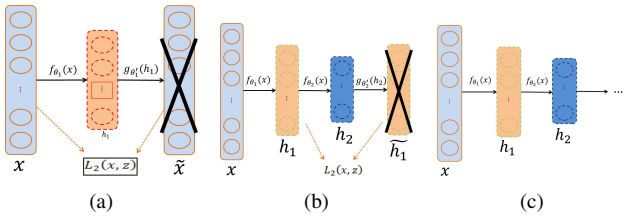
For traditional IQA, V1 area is the most important part for the general perception distortions. Specially, V1 is responsible for the noise and blur sensing. So the simple simulation on V1 will lead to wonderful results for traditional IQA. For image retargeting quality assessment, the problem is changing and special. In simple terms, finding the keys for IRQA is a different issue with IQA.

In order to better understand the various methods of image retargeting, we select some typical methods and make related transformations. The results are shown in Fig.1. It should be noted that the reference image is shown in Fig.1a. On this basis, the purpose is to obtain a retargeted image, in which the number of pixels in the horizontal direction is reduced to half. In this paper, eight operators are considered: Simple Scaling Operator (SSO) in Fig.1b, Manual Cropping (MC) in Fig.1c, Seam Carving (SC) in Fig.1d [3], Nonhomogeneous Warping (WARP) in Fig.1e [5], Scale and Stretch (SCST) in Fig.1f [6], Multi Operator (MULTI) in Fig.1g [7], Shift Map (SM) in Fig.1h [8] and Streaming video (STVI) in Fig.1i [9].

For newly IRQA, V4 area will be the critical part, which can be responsible in visual cognation and visual attention. In other word, image retargeting considers more in the geometry shape and content matching [28], [29]. According to the above analysis, many special features are important for IRQA based on V4 area.

In this paper, four kinds of geometry shape descriptors are used: local binary pattern, gradient map, difference of Gaussian, scale-invariant feature transform. In Section III-A, the details will be given. In addition, two kinds of content matching descriptors are used: learning similarity-preserving binary code and spatial envelope. In Section III-B, the two descriptors will be listed.

B. Motivation in Segmented Stacked AutoEnCoder for Image Retargeting Assessment



(a) (b) (c)

Fig. 2: Iterative training construction of the stacked AutoEnCoder (SAE) model. (a) Based on the training in first layer, reconstruction layer will be eliminated. (b) f_{θ_1} is defined as the encoding function in first layer and h_1 is defined as the training method for second layer. In this way, it can get the encoding function, f_{θ_2}, for the second layer. (c) Repeating the construction process.

The most simple way of deep learning is using the characteristics of artificial neural network, which has a hierarchical system structure. Given a neural network, we assume that the output and input are the same.Based on this assumption, then the deep network can train the parameters and make adjustment for weights of each layer. Naturally, we get several different representations of the input I , which can be regarded as feature. Autocoder is a neural network for repetition of the input signal [30], [31]. Inspired by the concept of good representation, stacked auto-encoders(SAE) was proposed [32], [33], which is shown in Fig.2.

The artificial neural network itself is a hierarchical structure. If a neural network is given, we assume that its output should be the same as the input, and then train the parameters of the neural network to obtain the weights in each layer. Naturally, we get several different representations of the input I, which are features. An automatic encoder is a neural network that can reproduce the input signal as much as possible. In order to achieve this replication, the automatic encoder must capture the most important factor that can represent the input data and find the principal components that can represent the original information.

In the process of sparse coding, the concept of the basis is very important. $O = a_1\Phi_1 + a_2\Phi_2 + ... + a_n\Phi_n$, where Φ_i is the base and a_i is the coefficient. Based on it, an optimization problem can be obtained. By constantly minimizing the difference between input and output, we obtain bases Φ_i and coefficients a_i. These bases and coefficients are another approximation of the input in Equ.1.

$$x = \sum_{i=1}^{k} a_i \phi_i \tag{1}$$

This process can also be learned automatically. If we add the L_1 regularity limit, we can get another form of expression in Equ.2.

$$Min|I - O| + u * (|a_1 + a_2 + ... + a_n|) \tag{2}$$

This method is called as sparse coding. In this way, a signal can be represented as a linear combination of bases. And it requires only a few bases to represent the signal. Through sparse components, we can represent input data. In this way, it will be helpful for different kinds of signal processing, especially for natural images. Images can be expressed as a superposition of basic elements, such as local surface or line.

Specifically, sparse coding can be divided into two parts. One is the training phase, and the other one is the coding phase. Given a series of sample images $[x_1, x_2, ...]$, we need to get a set of bases by learning, which is dictionary. The training phase is a process of repeated iterations. As mentioned above, we alternately change a and ϕ, making the following objective function minimum in Equ.3,

$$\min_{a,\phi} \sum_{i=1}^{m} \left\| x_i - \sum_{j=1}^{k} a_{i,j} \phi_j \right\|^2 + \lambda \sum_{i=1}^{m} \sum_{j=1}^{k} |a_{i,j}| \tag{3}$$

By fixing the dictionary ϕ_k, and then we can adjust the a_k to minimize the object function. Next, by fixing the a_k, we can then adjust the ϕ_k. Repeated iterations will be done until the convergence, which can give a dictionary choice.

Given a new image x, the sparse vector a can be obtained by solving a LASSO problem by the dictionary obtained above. This sparse vector is a sparse representation of the input vector x.

C. Reducing the Effect of Overfitting

Based on the above considerations, segmented stacked AutoEnCoder will be in good performance for IRQA. However, most of the deep learning methods will encounter the overfitting problem, especially when there are no enough samples in the training subset. In IRQA, the assessment processing will be based on only hundreds of training samples, which will be in trouble if we simply apply deep learning in this model. Generally, different kinds of regularization methods are considered: L1/L2 regularization (weight decay), data augmentation, early stopping and dropout.

When the objective function or cost function is optimized, a regular term can be added as regularization. The L_1 regularization is based on L_1 norm. In other word, L_1 norm of the parameter is added on the objective function in Equ.4:

$$C = C_0 + \frac{\lambda}{n} \sum_{\omega} |\omega| \tag{4}$$

Where C_0 represents the original cost function, n is the number of samples, and λ is the regular term coefficient, which weighs the proportion of λ and C_0. Specially, the latter item is the L_1 regular term [34], [35].

L_1 regularization is to make those parameters ($\omega \approx 0$) near to zero, so that more parameters will be zero, thus reducing the complexity of the model, in order to prevent overfitting and improve the generalization ability of the model [36]. Of course, more complex regularization methods will be applied in this approach to minimize the effect of overfitting.

When we use samples to train model or use this model to fit the future samples, the hypothesis is that the training data is independent, which is same as the testing data [37]. As a result, more samples will lead to better deep learning models. But we often don't have enough samples to be used. For example, some experiments require manual sample marking, resulting in inefficiencies and errors. At this point, we need to take some computational methods to operate on existing data sets to get more data.

In this paper, we have limited image retargeting results with subjective DMOS in the existing databases. So we must try methods to make use of the existing samples, such as regularization based on dropout.

III. Proposed method

The whole framework for the proposed method in this paper is shown in Fig.3. The special process of the features extraction and training phase will be given in details.

Different from traditional workflow for IRQA, the framework based on deep learning requires more low-level features input for the deep network. During our previous research process on image retargeting, we tried to use all high-level features for the framework. However, the performances are not satisfied. So we think that the IRQA is a special problem, which is very sensitive to the foundational information in retargeted images. So the traditional features such as LBP, GM and SIFT can be used for image representation. In essence, detecting the change of image shape and content information based on low-level features can better evaluate the effect of image retargeting.

Based on the above analysis, we evaluate the retargeted image on two directions. One is geometry shape changing extent compared with the referenced image, and the other one is content maintaining quality during the retargeting process. So the different evaluation directions require different features for state-of-the-art assessment performance.

After features extraction, segmented stacked AutoEnCoder can be used to simulate the final perception step in human vision system. And the detailed structure is shown in Sec.III-C. In addition, regularization based on dropout can overcome the overfitting problem in IRQA. Finally, final evaluation results can be obtained by combination of the two segmented stacked AutoEnCoders.

A. Prior Feature Descriptor for Geometry Shape

1) Local binary pattern: Local binary pattern (LBP) is a powerful descriptor to represent the marginal shape for images and it can also make texture classification [38]. First of all, we compare each pixel with its eight neighbors to construct LBP

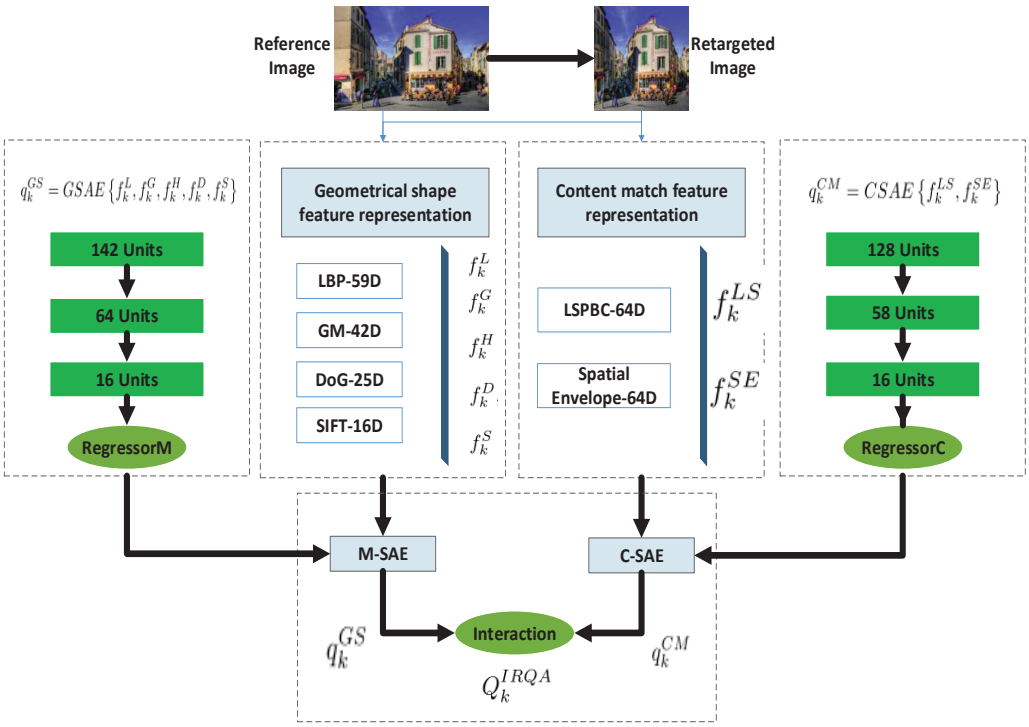


Fig. 3: The whole framework for the proposed method

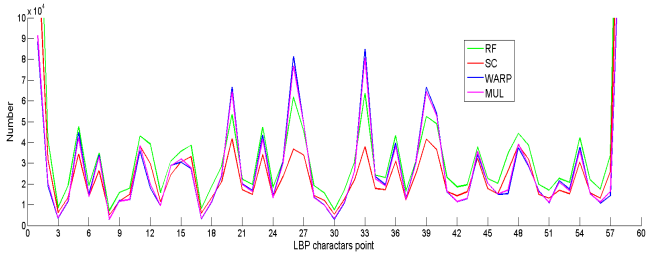


Fig. 4: Local binary pattern detection: green line represents the LBP for the referenced image, and the other three are for the different retargeted images

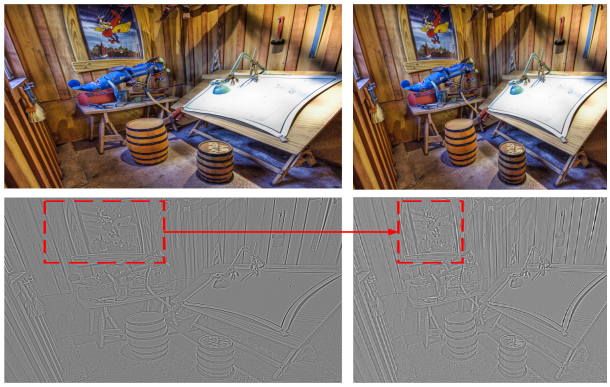


Fig. 5: Difference of Gaussian (DoG) for referenced image and retargeted image

descriptor and obtain the statistical results in 59 categories, which is shown in Fig.4. Then it can be used to represent the texture invariance. In this way, the LBP map can be obtained by the ratios between referenced images and retargeted images. It is important to emphasize that the LBP here uses a uniform pattern. As a result, that is a total of 59-D LBP feature descriptor for each retargeted image.

2) Gradient Map: In [39], Xue *et al.* used gradient map to make assessment for images. Inspired by this idea, we consider it for the IRQA. Specially, it can be computed in Equ.5.

$$G_I = \sqrt{(I * h_x)^2 + (I * h_x)^2} \tag{5}$$

where h_x and h_y are the gradient operators in both horizontal and vertical directions.

$$h_d(x, y|d) = -\frac{1}{2\pi\sigma^2}\frac{d}{\sigma^2}exp(-\frac{x^2+y^2}{2\sigma^2}), d \in \{x, y\} \tag{6}$$

Two scales will be used at implementation based on Equ.6. Then 16 AGGD fitting parameters and 2 GGD fitting parameters can be computed for each scale [40]. In addition, the magnitude, variance and entropy should be computed for each scale. In this way, a total 42-D GM feature descriptor can be obtained based on the ratios between retargeted image and referenced image [41].

3) Difference of Gaussian: Difference of Gaussian (DoG) is another effective Geometric features, which has been proved. We already know the low-pass filtering results by convolving

the image with the Gauss function can be used for the denoising process. The Gauss low-pass filter here is a function of the normal distribution. And the processing results are shown in Fig.5.

$$f(u,v,\sigma) = \frac{1}{2\pi\sigma^2 e^{-(u^2+v^2)/(2\sigma^2)}} - \frac{1}{2\pi K^2\sigma^2 e^{-(u^2+v^2)/(2K^2\sigma^2)}} \tag{7}$$

Subtraction results between two images in different parameters of the Gauss filter can be obtained based on Equ.7. In this way, we get the DoG diagram to represent the retargeted images or referenced images. In this paper, we set five different σ and the GGD fitting for it will get 10 parameters [42]. In addition, we can also compute the magnitude, variance and entropy for each scale, which can get 15 parameters [43]. In total, there are 25-D DoG feature descriptor based on the ratios between retargeted image and referenced image.

4) Scale-Invariant Feature Transform: Scale-invariant feature transform(SIFT) is a local feature detection algorithm as Equ.8, which can be used to find the interest points or corner points. In the process of image retargeting, the reservation of key points has a great influence on the perception, so it can be used as an important index to evaluate the retargeting results.

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \tag{8}$$

In this paper, we use the SIFT match as the feature points, which is shown in Fig.6. In the matching process, 16 different thresholds are set in the retrageted image. Therefore, there are 16-D SIFT feature descriptor based on the ratios between retargeted image and referenced image.

B. Prior Feature Descriptor for Content Matching

1) Learning Similarity-Preserving Binary Code: Learning similarity-preserving binary code was proposed for efficient similarity search in large-scale image collections [44], [45]. Inspired by the idea, we put forward the designed LSPBC for content match. In this way, the retargeting image can be coded as binary code, which can be used to compute the similarity between retargeted one and the referenced one.

In this paper, we use linear dimensionality reduction at first. Based on the resulting space, the binary quantization can be performed. In order to express in convenient, we have a set of n data points $\{x_1, x_2, ..., x_n\}, x_i \in \mathbb{R}^d$, which are the rows data for the image matrix $X \in \mathbb{R}^{n\times d}$. For the first step, the objective function in Equ.9 can be used for maximizing.

$$\Gamma(W) = \sum_k var\left(h(x)\right) = \sum_k var\left(sgn(x\omega_k)\right) \tag{9}$$

The variance can be maximized by the encoding functions. However, it can not meet the requirement for exact balancedness. The Equ.10 and 11 should be considered.

$$\Gamma(W) = \sum_k E(\|x\omega_k\|_2^2) = \frac{1}{n}\sum_k \omega_k^T X^T X \omega_k \tag{10}$$

Fig. 6: Scale-invariant feature transform match for retargeted image and referenced image

$$\Gamma(W) = \frac{1}{n} tr\left(W^T X^T X W\right), W^T W = I \tag{11}$$

If we assume $c \in \mathbb{R}^c$ as the vector in projected space, it is obvious that $sgn(v)$ can be as the vertex of the hypercube $(-1,1)^c$, which is closest to v. In this way, we should get the smallest quantization loss $\|sgn(v) - v\|^2$ for the resulting binary code to preserve the original local structure of the images.

Specially, it is an efficient alternating minimization scheme to find the rotation of zerocentered data, thus to minimize the quantization error of the two tested image. In other word, if the retargeting image is more consistent with the reference one in content match, the distance between learning similarity-preserving binary code will be smaller. However, the direct computation can not get a good result, which will be discussed in Section.V-D. So we set the obtained binary codes as the training data for the proposed segmented stacked AutoEncoder.

Through the LSPBCode computation, we evaluate code sizes up to 256 bits. In this paper, we take 8 bits as a feature value for each referenced image and retargeted image. Therefore there are 64-D LSPBC feature descriptor.

2) Spatial Envelope: In addition, Spatial Envelope was proposed for the recognition of real world scenes, which is a very low dimensional representation [46]. In Spatial Envelope, the different dimensions such as naturalness, openness, roughness, expansion and ruggedness can be extracted. Using the spatial envelope differences of the referenced image and the retargeted image, we can measure the degree of variation between the two which are shown in Fig.7.

The estimation for the spatial envelope can be done in different regression techniques. In this paper, image s from global spectral features v is defined as Equ.12:

$$\begin{aligned} \hat{s} &= \mathbf{v^T d} \\ &= \sum^{N_G} v_i d_i = \iint A(f_x,f_x)^2 DST(f_x,f_x) df_x df_y \end{aligned} \tag{12}$$

This can provide a simple interpretation for the representation. The discriminant spectral template(DST) can describe

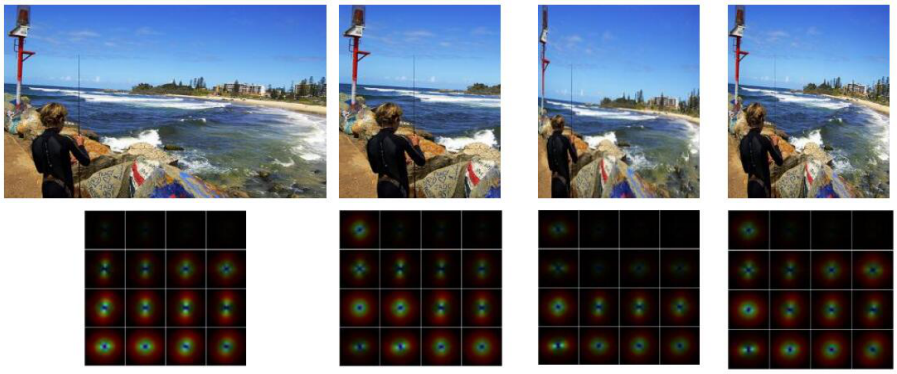


Fig. 7: Spatial Envelope description for referenced image and retargeted image

how each spectral component contributes to a spatial envelope property. The the similar estimation can be performed by spectrogram features.

Through the Spatial Envelope, 512 parameters will be obtained for each image. Then we make a two-para fitting based on the 16 groups for each referenced image and retargeted image. Then there are also 64-D Spatial Envelope feature descriptor.

C. Training Data Considering Segmented Stacked AutoEncoder

In [47], SAE is proved to outperform DBN in a particular case. As illustrated in Fig.3, the training model used in this paper contains two segmented 2D-SAEs. And the 2D-SAEs are with three hidden layer, which has been proved to be valid in related research studies. SAE is a good method for dealing with varying functions, which is consistent with the purpose in this paper. In addition, the deep structure of SAE can make a stronger learning than the shallow neural networks. Specifically, the two stages divided into training and testing should be explained respectively.

In the perception process of retargeted image, both the simple geometrical changes and complex content maintaining should be considered. For some retargeted images, geometrical shape changes are not obvious while content has been damaged during the retargeting. For others, the situations are on the contrary. If we use a whole AutoEnCoder as the regression method, the combination on two kinds of features before inputting will mislead the final assessment results.

Based on the above consideration, we propose the segmented stacked AutoEnCoder. Actually, it can be regarded as a two-branch Stacked AutoEnCoder, but the inner structures of the two are totally different. The one designed for GS assessment is called GSAE with 142-64-16 units and the other one designed for CM is called CSAE with 128-58-16 units. In this way, the segmented stacked AutoEnCoder can work better than the whole one.

In the unsupervised pre-training phase, there are three main parameters should be selected: learning rate, epoch number and the batch size. Through test, we set the learning rate as 0.5 for each hidden layer and the epoch number as 1000 in order to get the convincing results. It is particularly important to note that our batch size is 1. In other words, we adjust the training network based on every single input.

In addition, the full-batch train error is set as 0.005 for the first layer, and it will be changed as 0.001 for the other two layers in order to speed up the learning process.

Then we optimize the network by minimizing $J(\theta, b)$ as the final regression model structure.

$$J(\theta, b) = \frac{1}{2m}[\sum_{i=1}^{m}(h_{\theta,b(x^{(i)})-y^{(i)}})^2] \tag{13}$$

Specially, θ and b in Equ.13 is used to define the whole vector which contains all the weight values and bias values in the deep network. In addition, $J(\theta, b)$ can be regarded as the cost function. Through repeated data training, we can get the best Stacked AutoEnCoder framework based on the constraints of cost function, $J(\theta, b)$. And the weight values and bias values are optimized in this process. It should be emphasized that the SAE used in this paper comes is based on the Deep Learning Toolbox.

D. Regularization Design for Overfitting Problem

The algorithm proposed in this paper is designed for IRQA. However, The training data sets for IRQA are usually small, so we encountered the problem of overfitting when designing the deep network. Therefore, the regularization method is necessary, which is consistent with the content in Section II-C. Generally, different kinds of regularization methods are considered: L1/L2 regularization (weight decay), data augmentation, early stopping and dropout. Because L1/L2 regularizations work mainly through cost function, it is not suitable for deep network in this method. At present, there is no more reliable data that can be used for IRQA, so data augmentation is also not easy to achieve. In addition, we have also tried the early stopping method. At the end of every epoch, we calculate the accuracy of validation data, and stop training when accuracy is no longer improving. But it will bring more complexity to this problem and cannot effectively solve the overfitting problem in this research. For the specific situations of IRQA, too much network parameters will reduce the generalization robustness of the final network. So dropout is more suitable in this problem.

Based on the above consideration, we choose dropout as the final approach to decrease the effect of overfitting, and Equ.14 gives the example between l^{th} and $l+1^{th}$ layers.

$$y_{l+1}^{i} = f(\theta_{(l+1)}^{i}(r^{l} * y^{l}) + b_{(l+1)}^{i}), r_{(l)}^{i} \sim Bernoulli(p) \tag{14}$$

In Euq.14, $r_{(l)}^{i}$ can be regarded as the dropout factor, which can directly affect y^{l}. At the beginning of training, we set $p = 0.5$, which means that half of hidden layer neurons are randomly removed. Then it can solve the overfitting problem. By employing the regularization method, we impose additional constraints that indirectly reduce the number of parameters of the free variables.

E. Image Quality Pooling Based on Weighting Schemes

In the quality pooling phase, the GS features and CM features will be used in two different SAE structure respectively, which can get respective assessment results. The two obtained results can be combined according to Equ.16. So the deep model cannot be defined as end-to-end fashion. On the contrary, the features are still based on "hand-crafted" function, which is suitable for the requirement of IRQA.

Based on the Section I and II, the image retargeting quality is mainly affected by two main kinds of factors. The first category is the fundamental physical information change, which is called geometry change in this paper. The second category is content level based on semantic understanding, which is called content match in this paper. The two kinds of features are usually contradicted with each other. So we choose generate GS or CM respectively, and then combine them together, which can bring more precious quality assessment results than the methods based on end-to-end fashion.

In the quality pooling phase, the different features will be used in two different SAE structure respectively. By feeding the features into the fine-tuned SAE model, the quality scores based on marginal shape and content match will be get respectively.

$$q_{k}^{GS} = GSAE\left\{f_{k}^{L}, f_{k}^{G}, f_{k}^{H}, f_{k}^{D}, f_{k}^{S}\right\} \tag{15}$$

In Equ.15, q_{k}^{GS} is the image retargeting quality considering only geometric shape. In addition, $f_{k}^{L}, f_{k}^{G}, f_{k}^{H}, f_{k}^{D}, f_{k}^{S}$ is the features vectors for local binary pattern, gradient map, histogram of oriented gradient, difference of Gaussian and SIFT descriptor respectively.

Using the same principle, we can get the image retargeting quality considering only content match, q_{k}^{CM}. In Eq.16, f_{k}^{LS} represent the feature vectors based on learning similarity-preserving binary code. The f_{k}^{SE} can be regarded as the feature characters computed by Spatial Envelope.

$$q_{k}^{CM} = CSAE\left\{f_{k}^{LS}, f_{k}^{SE}\right\} \tag{16}$$

However, the quality scores based on the last two parts cannot represent the final image retargeting quality perfectly. In order to solve this problem, we introduce the weighting schemes to get the final score. By combining the q_{k}^{GS} and q_{k}^{CM}, we can get the final IRQA score, Q_{k}^{IRQA}.

$$Q_{k}^{IRQA} = (q_{k}^{GS})^{\omega} \cdot (q_{k}^{CM})^{(1-\omega)} \tag{17}$$

After many experiments, we get a best proportion index in Equ.17, that is $\omega = 0.423$. Such a proportion setting can better utilize the geometrical and contextual interaction for the IRQA.

IV. Benchmark Databases and Performance Protocol

In this section, three benchmark databases and the performance protocol are introduced for further experiments.

A. Benchmark Databases

In order to better validate the effectiveness of the proposed algorithm, three well-known retargeting image assessment databases are used, which are RetargetMe [21], CUHK [22] and NRID [23]. In these databases, subjective evaluation results are added. Three databases are presented with details below.

1) RetargetMe: In [21], the RetargetMe database was first proposed for the image retargeting quality assessment. As a benchmark database, 37 source images are contained in RetargetMe. For every source image, eight retargeting methods are used as the operators, which has been introduced in Section.II-A. In this way, there are total 296 retargeted images are generated. According to [48], [49], the paired comparison manner is used in subjective assessment. Specially, subjects should choose the better one in every retargeting pairs based on the same source image. At last, the number of chosen times for the retargeted result can be regarded as the subjective rating score.

2) CUHK [22]: There are 57 source images in it and three retrageting operators are used for each one. It is worth mentioning that this database used a total of ten retargeting methods, and each original image selected three randomly. Specially, eight retargeting operators used in RetargetMe were also used in CUHK. And two other methods were added, which can be called optimized seam carving [50] and energy based deformation [51]. In this way, 171 retargeted images are operated as the results. In subjective assessment, a five-category discrete quality scale model was used in CUHK, which is different with RetargetMe database. At last, the final mean opinion score(MOS) can recorded as the subjective results.

3) NRID [23]: In NRID, 35 source images are contained, and five main retargeted operators are used, which is less than the last two database. Specially, Seam Carving (SC) [3], Nonhomogeneous Warping (WARP) [5], Multi Operator (MULTI) [7], Simple Scaling Operator (SSO), and Shift Map (SM) [8] are included. This database can test the accuracy of the algorithm on the mainstream method designed for image retargeting more accurately. Naturaly, 175 retargeting image results can be operated for NRID. At last, the subjective assessment is made according to the the paired comparison manner [48], which is same as RetargetMe database.

B. Performance Protocol

There are two reasons for using different evaluation measure criteria in different databases. On the one hand, it is corresponding with the general protocol design in the field of image retargeting assessment. So we can compare the proposed methods with other in this way. On the other hand, the different performance protocols are suitable for different database building principles. RetargetedMe database and NRID database are based on paired comparison methodology, so we choose Kendall tau as performance protocol for both of them. However, CUHK database is built based on ACR methodology, we choose PLCC, SRCC, RSME and OR as the performance protocol for it.

For RetargetMe database and NRID database, the correlations evaluation between objective and subjective scores can be measured as Kendall τ [21]:

$$Kendall\tau = \frac{n_c - n_d}{0.5n(n-1))} \tag{18}$$

In Eq.18, n is the ranking length, n_c is the concordant pairs number, and n_d is the discordant pairs number.

For CUHK database, four evaluation metrics as traditional image quality assessment are used to evaluate the correlations between objective and subjective scores: PLCC(Pearson Linear Correlation Coefficient), SRCC(Spearman Rank-order Correlation Coefficient), RMSE(Root Mean Squared Error) and OR(Outlier Ratio) [52].

PLCC(Pearson Linear Correlation Coefficient) can be computed with nonlinear regression, and the regression process can be made by Eq.19, which was proposed by Sheikh *et al.* [53].

$$f(x) = \beta_1(\frac{1}{2} - \frac{1}{1+e^{\beta_2(x-\beta_3)}}) + \beta_4 x + \beta_5 \tag{19}$$

SRCC(Spearman Rank-order Correlation Coefficient) can be used to measure the monotonicity for the objective image retargeting quality assessement. The third measure metric is the RMSE(Root Mean Squared Error), which can be computed between the subject scores and the objective scores after the nonlinear regression. The last one is OR(Outlier Ratio) [52], which can reflect the ratio between the false objective score and the total score number. Specially, the false score can be defined as the one which lies outside of the $[MOS-2\sigma, MOS+2\sigma]$ interval, where σ is the corresponding standard deviation.

Larger PLCC and SRCC values indicate that the objective evaluation value is better consistent with the subjective evaluation value. And Smaller RMSE and OR values can indicate the good results for the IRQA.

V. Experimental Results

In this section, the experimental results will be given. On the one hand, the results of many recognized algorithms designed for IRQA will be enumerated, and their effectiveness will be analyzed. On the other hand, the method designed in this paper will also be applied to the same situation for comparison purposes. In addition, cross database experiments are also made, and this part of the results will be analyzed in detail. What's more, the algorithm framework presented in this paper is complex and contains more variable parameters. So, how the changes in framework affect experimental results is also needed to be considered. At last, advantages and disadvantages need to be discussed.

TABLE I: Overall performances on RetargetMe database and NRID database

Mertic	RetargetMe		NRID	
	Kendall τ	STD	Kendall τ	STD
BDS [17]	0.083	0.268	0.131	0.527
EMD [54]	0.251	0.272	0.362	0.361
SIFT-flow [55]	0.145	0.262	-0.011	0.502
EH [56]	0.004	0.334	0.108	0.556
CSIM [18]	0.182	0.258	0.154	0.512
SR [24]	0.413	0.282	0.577	**0.334**
proposed	**0.476**	**0.243**	**0.598**	0.412

TABLE III: Overall performances on CUHK database.

Mertic	PLCC	SRCC	RMSE	OR
BDS [17]	0.2896	0.2887	12.922	0.2164
EMD [54]	0.2760	0.2904	12.977	0.1696
SIFT-FLOW [55]	0.3141	0.2899	12.817	0.1462
EH [56]	0.3422	0.3288	12.686	0.2047
CSIM [18]	0.4374	0.4662	12.141	0.1520
PGDIL [23]	0.4622	0.4760	10.932	0.1345
ARS [20]	0.6835	0.6693	9.855	0.0702
Proposed	**0.7012**	**0.6732**	**8.364**	**0.0574**

A. Performance Comparisons on RetargetMe Database and NRID Database

In Table.I, the average rank correlation experimental results in RetargetMe database and NRID database are shown respectively. Specially, the Kendall τ distance and the standard deviation based on different IRQA methods are given. From the table, we can see that the predicted objective results are more consistent with the subjective ranking. In RetargetMe database, the mean Kendall τ distance is larger than 0.47. And it is also larger than 0.59 in NRID database. The proposed method is compared with BDS [17], EMD [54], SIFT-flow [55], EH [56], CSIM [18] and SR [24]. Both of the experimental results are the best in the compared methods.

We analyze the main reason for the good performance may be that the proposed deep framework can train a more robust learning structure. Compared with the method of shallow learning, the proposed method has a deeper excavation of the feature descriptors.

In addition, the performances for different image subsets in RetargetMe with labelled attributes are also shown as good results. It can achieve the best correlations compared with other IRQA methods, which is shown in Table.II.

B. Performance Comparisons on CUHK

In order to verify the effectiveness of the proposed IRQA algorithm, we calculate the rank correlation results on the

TABLE II: The performances for different image subsets in RetargetMe with labelled attributes (The best results are highlighted)

Mertic	Kendall τ in each subset						Total RetargetMe	
	Line Edge	Faces people	Foreground	Texture	Geometric structure	Symmetry	Kendall τ	STD
BDS [17]	0.040	0.190	0.167	0.060	-0.004	-0.012	0.083	0.268
EMD [54]	0.220	0.262	0.226	0.107	0.237	0.500	0.251	0.272
SIFT-flow [55]	0.097	0.252	0.218	0.161	0.085	0.071	0.145	0.262
EH [56]	0.043	-0.076	-0.079	-0.060	0.103	0.298	0.004	0.334
IR-SSIM [19]	0.309	0.452	0.277	0.321	0.313	0.333	0.363	0.271
ARS [20]	0.463	**0.519**	0.444	0.330	0.505	**0.464**	0.452	0.283
Proposed	**0.466**	0.512	**0.452**	**0.434**	**0.515**	0.443	**0.476**	**0.243**

benchmark CUHK database, and the comparison results are shown in Table.III. The proposed method is compared with BDS [17], EMD [54], SIFT-FLOW [55], EH [56], CSIM [18], PGDIL [23] and ARS [20].

In the performance comparisons on CUHK, we give the mean and standard deviation values for rank correlations in Table.III. In the comparison of the results, we can see that the method proposed in this paper can get better performance result than the known state-of-the-art methods.

TABLE IV: Cross databases performance

Training database	Testing database	PLCC	SRCC
RetargetMe	CUHK	0.452	0.224
	NRID	0.412	0.243
CUHK	RetargetMe	0.465	0.312
	NRID	0.532	0.314
NRID	RetargetMe	0.422	0.246
	CUHK	0.463	0.247

C. Cross Database Performance

In the application process, different image resources need to be processed, and the results of cross database experiments are very important to this issue. However, most of the IRQA methods did not give the cross database performance. In this paper, we design the cross database experiments based on the three IRQA databases which are mentioned above, RetargetMe, CUHK and NRID. Specially, we train it on one retargeted image database and test it based on the other one. It is necessary to point out that the score standards in three database are different, so we give RetargetedMe and NRID for the scores as CUHK according to the preferred number.

For RetargetedMe and NRID, we set virtual DMOS for quality of every retargeted image based on the preferred number in original paired comparison methodology. Specially, the adjustment is based on linear regression, and it can change the pair preferred number into DMOS between 0 and 100 for every retargeting image, which corresponds to CUHK.

With the transaction, we can guarantee that the three databases can have the same score standard. In this model, we can use PLCC and KRCC to measure cross library evaluation algorithms. Table.IV shows the cross database performance. Of course, such a result is worse than testing in a single database alone. However, it shows that proposed method can achieve certain performance in cross database experiment, and it has certain practical value.

D. Impact of Each Framework Component

1) Feature Extraction : The extracted features are important for the image representation. In this paper, the features can be divided into two categories: Geometrical shape and content matching. So we want to detect how the different feature parts affect the final evaluation results. Specifically, we did two parts of the experiment.

On the one hand, we use one of the feature categories one by one to obtain the evaluation results. In this way, we can see the role of a single feature category. Table.V shows the experimental results in this way. On the other hand, we remove the feature category one by one and use other features to obtain the evaluation results. In this way, we can see the experimental results after the lack of a single feature category, which are also shown in Table.V.

Based on the experimental results, we can make a evaluation on different feature categories. Specially, the feature extraction based on spatial envelope is better than others for image retargeting quality assessment. For the single test by content matching, its evaluation result is better than the test by geometric features.

2) Segmented Stacked AutoEncoder: Prior to this, many researchers applied deep learning to traditional image quality evaluation or stereo image quality evaluation. For example, Shao *et al.* [57] applies depth belief network to stereo image quality evaluation, and has achieved very good results.

As the main contribution, the segmented stacked AutoEnCoder based IRQA was proposed in this paper. But we need quantitative measurements of how depth learning can improve IRQA's capabilities. In this section, we did experiments using SAR instead of SAE. At the same time, DBN is also used instead of SAE to verify its effectiveness. Table.VI show the experimental results in this way.

In addition, the parameters in SAE are important for the performance. Specifically, the hidden units number in the two segmented stacked AutoEnCoder frameworks will affect the experimental results seriously. So we change the two parameters and test the system performance in the three databases. And the results are shown in Fig.8. According to the experimental results, we can draw the conclusion that the hidden units number for the two segmented stacked AutoEnCoder should be 64 and 58 perspectively.

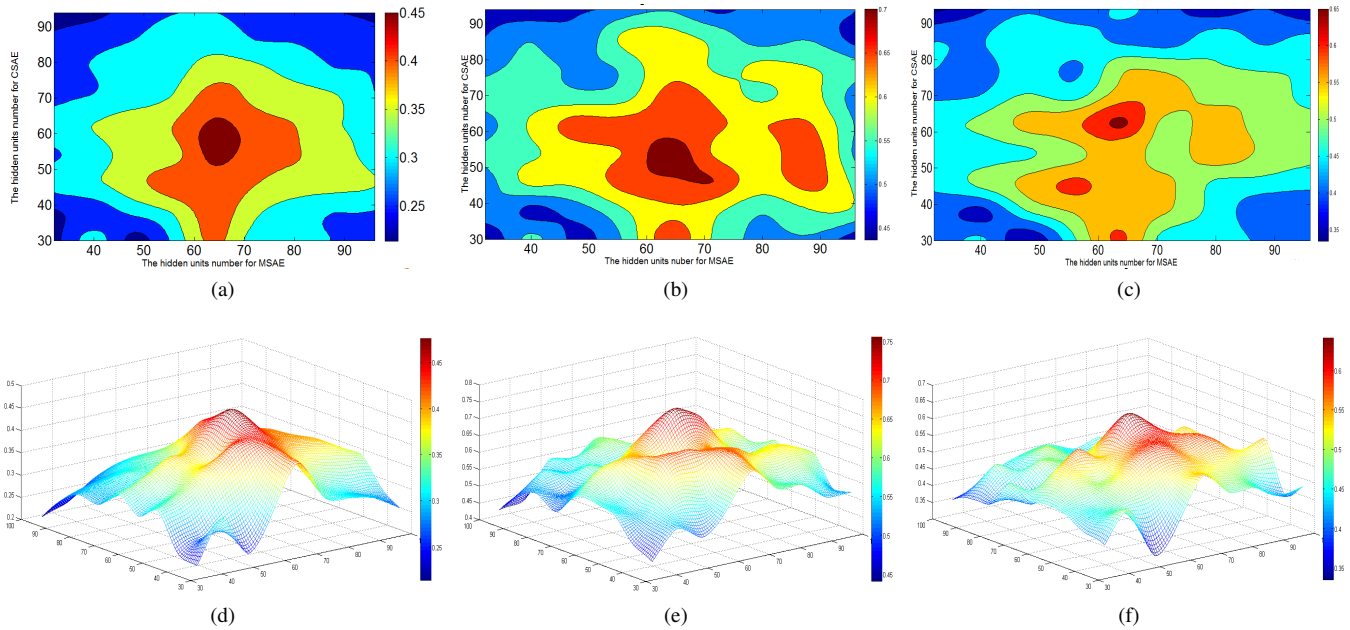


Fig. 8: The over performance changes in three databases based on the hidden units number in the two segmented stacked AutoEnCoder frameworks (a) and (d) are performances in RetargetedMe; (b) and (e) are performances in CUHK; (c) and (f) are performances in NRID.

TABLE V: The performances for different image descriptors in RetargetMe and NRID database

Only with the special feature		f_k^L	f_k^G	f_k^D	f_k^S	f_k^{LS}	f_k^{SE}	f_k^{GS}	f_k^{CM}	Overall
RetargetMe	Kendall τ	0.243	0.291	0.267	0.114	0.312	0.283	0.374	0.403	0.476
	STD	0.640	0.790	0.432	0.304	0.527	0.134	0.268	0.183	0.243
NRID	Kendall τ	0.234	0.247	0.115	0.260	0.334	0.363	0.511	0.524	0.598
	STD	0.135	0.325	0.241	0.316	0.235	0.341	0.246	0.312	0.412
Only without the special feature		$-f_k^L$	$-f_k^G$	$-f_k^D$	$-f_k^S$	$-f_k^{LS}$	$-f_k^{SE}$	$-f_k^{GS}$	$-f_k^{CM}$	Overall
RetargetMe	Kendall τ	0.392	0.390	0.362	0.304	0.352	0.383	0.403	0.374	0.476
	STD	0.342	0.123	0.146	0.147	0.234	0.353	0.183	0.268	0.243
NRID	Kendall τ	0.478	0.403	0.442	0.324	0.472	0.483	0.524	0.511	0.598
	STD	0.382	0.394	0.352	0.362	0.304	0.412	0.312	0.246	0.412

TABLE VI: Overall performances comparison for SVM, DBN and SAE based on RetargetMe database and NRID database.

	RetargetMe		NRID	
	Kendall τ	STD	Kendall τ	STD
SVM	0.359	0.368	0.431	0.452
DBN	0.402	0.372	0.462	**0.311**
SAE	**0.476**	**0.243**	**0.598**	0.412

E. Advantages and Limitations

1) Advantages: In this paper, we consider the IRQA based on segmented stacked AutoEnCoder. Prior to this, more IRQA methods choose using a shallow layer of neural networks for data training and testing , which are not corresponding with the perception of retargeted images in human vision system. The method of deep learning can better mine the internal relations between the extrated data, so as to achieve better evaluation results. In addition, by combining the score q_k^M based on marginal shape and the score q_k^C based on content match, the method in this paper can measure the retargeted image quality, which is more consistent with the actual perception feel.

2) Limitations: First of all, the use of depth learning methods for data training and testing are subject to overfitting problems. In the field of image evaluation, the number of samples that can be used is usually limited, which is usually only a few hundreds. The limited samples number will naturally cause us to suffer from a fitting problem. In this paper, we've worked hard to solve this issue by improving the network structure, which has been discussed in Section III-D. In future work, we hope to build a larger IRQA database to

avoid overfitting problems. In addition, the ultimate goal of this algorithm design is to promote the continuous progress of image retargeting technology. However, majority of current IRQA methods are seriously depending on the fixed database. In other words, more cross database tests should be done and compared to improve the application value for the IRQA.

VI. Conclusion

As a new research topic, image retargeting has been paid increasing attention. However, how to evaluate the results of different image retargeting is a very critical issue. In this paper, we propose a deep evaluator for image retargeting quality assessment. The overfitting problem caused by the number of samples has been solved by regularization. In order to make accurate evaluation for the retargeting results, we combine the geometry features and content features with two segmented stacked AutoEnCoders. Then the weighting schemes are used to combine the two scores. As the main contribution in this paper, the proposed method can simulate the retargeted image perception process in human vision system, and it can make accurate objective evaluation on the quality of experience. Experimental results based on three well known databases show that our method can evaluate different image retargeting results in superiority. Based on the designed algorithm, it can be used mainly into two application directions. The one is evaluating the different image retargeting results and assessing it whether can meet the playing requirement. The other one is to support further development of different image retargeting algorithms. For the first one, it is obvious and intuitive. The application for the second direction is more valuable. In addition, image retargeting quality assessment should be an interesting and promising future research direction to be explored.

References

[1] Y. Fang, Z. Fang, F. Yuan, Y. Yang, S. Yang, and N. N. Xiong, "Optimized multioperator image retargeting based on perceptual similarity measure," *IEEE Transactions on Systems Man and Cybernetics: Systems*, vol. 47, no. 11, pp. 2956–2966, 2017.

[2] Y. Xia, L. Zhang, R. Hong, L. Nie, Y. Yan, and L. Shao, "Perceptually guided photo retargeting," *IEEE Transactions on Cybernetics*, vol. 47, no. 3, pp. 566–578, 2016.

[3] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," in *ACM SIGGRAPH*, 2007, p. 10.

[4] J. Shen, D. Wang, and X. Li, "Depth-aware image seam carving." *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1453–1461, 2013.

[5] L. Wolf, M. Guttmann, and D. Cohenor, "Non-homogeneous content-driven video-retargeting," in *IEEE International Conference on Computer Vision*, 2007, pp. 1–6.

[6] 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, pp. 32–39, 2008.

[7] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *Acm Transactions on Graphics*, vol. 28, no. 3, pp. 1–11, 2013.

[8] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *IEEE International Conference on Computer Vision*, 2009, pp. 151–158.

[9] M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," in *ACM SIGGRAPH Asia*, 2009, p. 126.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[11] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[12] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[13] M. Narwaria and W. Lin, "Svd-based quality metric for image and video using machine learning," *IEEE Trans Syst Man Cybern B Cybern*, vol. 42, no. 2, pp. 347 – 364, 2012.

[14] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference video quality assessment of compressed video sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 10, pp. 1441–1456, 2012.

[15] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J. R. Ohm, and G. J. Sullivan, "Video quality evaluation methodology and verification testing of hevc compression performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 76–90, 2016.

[16] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranza, "An image quality evaluation method based on digital watermarking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 1, pp. 98–105, 2007.

[17] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[18] Y. Liu, X. Luo, Y. Xuan, W. Chen, and X. Fu, "Image retargeting quality assessment," in *Computer Graphics Forum*, 2011, pp. 583–592.

[19] 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 Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 4, no. 1, pp. 95–105, 2014.

[20] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration based aspect ratio similarity (ars) for image retargeting quality assessment." *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4286–4297, 2016.

[21] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," in *ACM SIGGRAPH Asia*, 2010, p. 160.

[22] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 626–639, 2012.

[23] 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 Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 377–389, 2014.

[24] Q. Jiang, S. Feng, W. Lin, and G. Jiang, "Learning sparse representation for objective image retargeting quality assessment," *IEEE Transactions on Cybernetics*, vol. 48, no. 4, pp. 1276–1289, 2017.

[25] L. Ma, L. Xu, Y. Zhang, Y. Yan, and K. N. Ngan, "No-reference retargeted image quality assessment based on pairwise rank learning," *IEEE Transactions on Multimedia*, vol. 18, no. 11, pp. 2228–2237, 2016.

[26] Y. Liang, Y. J. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms." *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 2, pp. 1099–1110, 2016.

[27] A. Liu, W. Lin, H. Chen, and P. Zhang, "Image retargeting quality assessment based on support vector regression," *Signal Processing Image Communication*, vol. 39, no. 2, pp. 444–456, 2015.

[28] Y. Wang, B. S. Peterson, and L. H. Staib, "Shape-based 3d surface correspondence using geodesics and local geometry," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. 644–651.

[29] R. Kwitt and A. Uhl, "Modeling the marginal distributions of complex wavelet coefficient magnitudes for the classification of zoom-endoscopy images," in *IEEE International Conference on Computer Vision*, 2007, pp. 1–8.

[30] J. Snoek, R. P. Adams, and H. Larochelle, "Nonparametric guidance of autoencoder representations using label information," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2567–2588, 2012.

[31] J. Deng, Z. Zhang, F. Eyben, and B. Schuller, "Autoencoder-based unsupervised domain adaptation for speech emotion recognition," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1068–1072, 2014.

[32] H. Larochelle, "Greedy layer-wise training of deep networks," *Advances in Neural Information Processing Systems*, vol. 19, pp. 153–160, 2007.

[33] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *International Conference on Machine Learning*, 2007, pp. 473–480.

[34] J. G. Park and S. Jo, "Approximate bayesian mlp regularization for regression in the presence of noise," *Neural Networks*, vol. 83, pp. 75–85, 2016.

[35] E. Castro, R. D. Hjelm, S. M. Plis, L. Dinh, J. A. Turner, and V. D. Calhoun, "Deep independence network analysis of structural brain imaging: Application to schizophrenia," *IEEE Transactions on Medical Imaging*, vol. 35, no. 7, pp. 1729–1740, 2016.
[36] J. Fan, T. Zhao, Z. Kuang, Y. Zheng, J. Zhang, J. Yu, and J. Peng, "Hd-mtl: Hierarchical deep multi-task learning for large-scale visual recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1923–1938, 2017.
[37] L. Liao, W. Jin, and R. Pavel, "Enhanced restricted boltzmann machine with prognosability regularization for prognostics and health assessment," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7076–7083, 2016.
[38] T. Ojala, "Gray scale and rotation invariant texture classification with local binary patterns," in *European Conference on Computer Vision*, 2002, pp. 404–420.
[39] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features." *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.
[40] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.
[41] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Making image quality assessment robust," in *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 1718–1722.
[42] Y. H. Lin and J. L. Wu, "Quality assessment of stereoscopic 3d image compression by binocular integration behaviors," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1527–1542, 2014.
[43] D. Tao, "Sparse representation for blind image quality assessment," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1146–1153.
[44] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2916–2929, 2013.
[45] Y. Gong and S. Lazebnik, "Iterative quantization: A procrustean approach to learning binary codes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 817–824.
[46] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
[47] M. Bianchini and F. Scarselli, "On the complexity of neural network classifiers: a comparison between shallow and deep architectures." *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 8, pp. 1553–1565, 2014.
[48] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3, pp. 324–345, 1952.
[49] R. A. Bradley, "Rank analysis of incomplete block designs: Ii. additional tables for the method of paired comparisons," *Biometrika*, vol. 41, no. 3, pp. 502–537, 1954.
[50] W. Dong, N. Zhou, J. C. Paul, and X. Zhang, "Optimized image resizing using seam carving and scaling," *Acm Transactions on Graphics*, vol. 28, no. 5, pp. 1–10, 2009.
[51] K. Z., F. D., and G. C., "Energybased image deformation," *Computer Graphics Forum*, vol. 28, no. 5, pp. 1257–1268, 2010.
[52] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.
[53] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms." *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.
[54] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *IEEE International Conference on Computer Vision*, 2010, pp. 460–467.
[55] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011.
[56] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, 2001.
[57] F. Shao, W. Tian, W. Lin, G. Jiang, and Q. Dai, "Towards a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions." *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2059–2074, 2016.

Bin Jiang received the B.S. and M.S. degree in communication and information engineering from Tianjin University, Tianjin, China, in 2013 and 2016. He is currently pursuing the Ph.D. degree at the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. He was also a visiting scholar in Ember-Riddle Aeronautical University, Daytona Beach, US. His research interests lie in image processing, including image quality assessment, stereo vision research, virtual reality and image security.

Jiachen Yang (M'13) received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, Tianjin, China, in 2005 and 2009, respectively. He is currently a professor at Tianjin University. He is also a visiting scholar with the Department of Computer Science, School of Science, Loughborough University, UK. His research interests include image quality evaluation, stereo vision research, pattern recognition and virtual reality.

Qinggang Meng (M'06-SM'18) received the B.S. and M.S. degrees from the School of Electronic Information Engineering, Tianjin University, China, and the Ph.D. degree in computer science from Aberystwyth University, U.K. He is a Reader with the Department of Computer Science, Loughborough University, UK. He is a fellow of the Higher Education Academy, UK. His research interests include biologically inspired learning algorithms and developmental robotics, service robotics, robot learning and adaptation, multi-UAV cooperation, human motion analysis and activity recognition, activity pattern detection, pattern recognition, artificial intelligence, and computer vision.

Baihua Li received her B.S. and M.S. degrees in Electronic Engineering from Tianjin University and her Ph.D. degree in Computer Science from Aberystwyth University, Aberystwyth, UK. She has worked at Tianjin University and Manchester Metropolitan University before she is now a Senior Lecturer in the Department of Computer Science at Loughborough University, UK. Her research emphasizes innovations and novel applications of machine learning, computer vision and pattern recognition techniques in various fields. More than 70 papers have been published in high impact journals and conferences of international standard, such as Pattern Recognition, the IEEE Transactions on Systems, Man, and Cybernetics and the IEEE Transactions on Biomedical Engineering.

Wen Lu (M'13) received the M.S. and Ph.D. degrees in electrical engineering from Xidian University, China, in 2006 and 2009, respectively. He was a Post-doctoral in Stanford University from 2010 to 2012. He is currently an Associate Professor at Xidian University, China. His research interests include image and video understanding, visual quality assessment, and computational vision.