Multi-Task Deep Learning for No-Reference Screen Content Image Quality Assessment

No Author Given

No Institute Given

Abstract. The past decades have witnessed growing development of image quality assessment (IQA) for natural images (NIs). However, since screen content images (SCIs) exhibit different visual characteristics from the NIs, few of NIs-oriented IQA methods can be directly applied on SCIs. In this paper, we present a quality prediction approach specially designed for SCIs, which is based on multi-task deep learning. First, we split a SCI into 32×32 patches and design a novel convolutional neural network (CNN) to predict the quality score of each SCI patch. Then, we propose an effective adaptive weighting algorithm for patch-level quality score aggregation. The proposed CNN is built on an end-to-end multi-task learning framework, which integrates the histogram of oriented gradient (HOG) features prediction task to the SCI quality prediction task for learning a better mapping between input SCI patch and its quality score. The proposed adaptive weighting algorithm for patch-level quality score aggregation further improves the representation ability of each SCI patch. Experimental results on two-largest SCI-oriented databases demonstrate that our proposed method is superior to the state-of-the-art no-reference IQA methods and most of the full-reference IQA methods.

Keywords: No-reference image quality assessment · screen content images · multi-task learning · histogram of oriented gradient.

1 Introduction

With the rapid development of various multimedia applications and social communication systems over the internet, screen content images (SCIs) have been widely introduced in people's daily life, such as online education, online browsen, remote screen sharing, etc. Undoubtedly, the visual quality of SCIs has a significant influence on viewing experience of the client side. Hence, it's highly desired to devise an effective image quality assessment (IQA) method aiming to automatically predict the objective quality of SCIs. However, as composite images, SCIs have significantly different properties compared to natural images (NIs). Thus the IQA models devised for NIs can not be directly employed in quality prediction of SCIs. Specifically, NIs containing natural scenes usually hold relatively smooth edges, complicated shapes and rich color with slow color change, while SCIs involve a mixture of sources which come down to natural content and computer-generated content (texts, charts, maps, graphics, symbols, etc.), especially contain plenty of text content, which results in SCIs have multiple sharp

edges, high contrast, relatively uncomplicated shapes and little color variance. Consequently, there is a strong need to design a SCIs-oriented IQA method. At present, the objective IQA methods for SCIs can be classified into three categories depending on the accessibility of reference images: full reference (FR), reduced reference (RR) and no reference (NR). The NR-IQA model is more practicable due to its fewer restrictions and broader application prospects, of which one is based on the hand-crafted features and regression model that is also referred to as a two-step framework, and the other is based on the CNN model, also called as end-to-end framework.

For the end-to-end framework, Zuo et al. [19] devised a novel classification network to train the distorted image patches for predicting the quality scores, and weights determined by gradient entropy were applied to fuse the quality scores of textual and pictorial images patches. In [3], a well designed CNN architecture was presented to evaluate the quality scores of small patches, and representative patches were selected according to the saliency map to accelerate the quality score aggregation for the SCIs. Chen et al. [2] proposed a naturalization module to transform IQA of NIs into IQA of SCIs. It's not hard to conclude that most of CNNs-based NR-IQA methods for SCIs split SCIs into patches aiming to acquire enough training data and simply assign the subjective quality score of an image to all the local patches as their local quality label. This is problematic because local perceptual quality is not well-defined and not always consistent with the image quality score. To partially represent the region-wise perceptual quality variation for SCIs, an adaptive weighting method is proposed to fuse the local quality for the quality of distorted image. However, most of existing methods focus on edge or gradient information as the important factor, even ignore the whole content information of SCI. Therefore, there is a strong desire for effective network structure designed for accurate quality prediction results and effective strategy of fusing local quality, which motivate us to propose an effective NR-IQA approach tailored for SCIs.

In this work, we suggest a novel CNN-based multi-task learning model specifically designed for SCIs. Now that the contents of SCIs are rich in texture, gradient and contrast information, considering that the human visual system (HVS) is highly sensitive to edge and texture information often encountered in SCIs, we design a histogram of oriented gradient (HOG) aided convolutional neural network to evaluate quality scores of SCI patches by exploring HOG features of SCIs as predictive information source, which is shortened as HOGAMTL. The main contributions lie in three aspects:

(1) A valid multi-task learning network is developed by taking the advantages of HOG features, which induces the CNN feature extractor to extract more texture features inherited from SCIs' contents. It achieves promising performance on benchmark databases and surpasses the compared methods.

(2) Considering influence of the different image patches' contents on the quality of the image, VLSD and local entropy are involved in acquiring the local weight of each patch. That ensures our model follows the characteristic of HVS.

(3) Unlike other adaptive weighting methods which simply regard the lo-

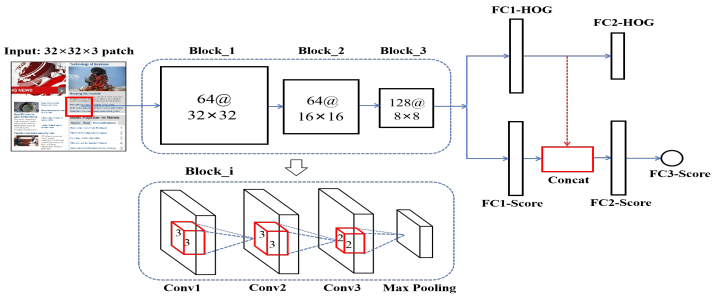

Fig. 1. An illustration of the architecture of our multi-task learning CNN model.

cal weight as the patch's weight for quality aggregation, our method embeds global weight in the estimation of each patch's weight, which further enhances the representation ability of each SCI patch.

2 The Proposed SCI-IQA Method

Fig. 1 illustrates the devised multi-task learning CNN architecture, which consists of two learning tasks, namely HOG features prediction task and quality score prediction task. The former assists the latter in training and learning a better map between the input image patch and its quality score. The overall pipeline of our approach is depicted into two parts. First, we detail our proposed CNN model, then we present a quality aggregation algorithm to fuse the local quality scores for obtaining a quality score of the image.

2.1 Patch-Level IQA Score Prediction

Image Preprocessing Regarding that CNN is sensitive to the mean value of the input data, the proposed CNN model for patch-level quality score prediction takes locally normalized SCI patches as input, which is defined as:

$$\hat{I}(i,j,d) = \frac{I(i,j,d) - \mu(i,j,d)}{\sigma(i,j,d) + C} \tag{1}$$

$$\mu(i,j,d) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} I_{k,l}(i,j,d) \tag{2}$$

$$\sigma(i,j,d) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} (I_{k,l}(i,j,d) - \mu(i,j,d))^2} \tag{3}$$

where $\omega = \{\omega_{k,l} | k = -K, \ldots, K, l = -L, \ldots, L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to three standard deviations and rescaled to a unit volume. $I(i,j,d)$ denotes the pixel value. K=L=3 mean the normalization window sizes, and C=1 is a constant that prevents division by zero.

Fig. 2. Examples of patch selection.

Network Training The architecture of our proposed CNN model for SCI patch quality score prediction is shown in Fig. 1. Corresponding to two prediction tasks, there are two outputs in this multi-learning network, which are 36 dimensions output of HOG features (denoted by FC_2-HOG) and one dimension output of SCI patch quality score (denoted by FC_3-Score). For the auxiliary HOG prediction task, considering the characteristic that SCIs contain more edge and texture information, we believe that HOG can always represent distortion situation of SCIs adequately and induce the CNN feature extractor to extract more texture features. The details about HOG extraction could refer to [4].

As a NR-IQA model, quality score prediction is the main task of our proposed HOGAMTL. It shares the CNN feature extractor with HOG features prediction task and concatenates feature vectors of two tasks to get a new feature vector for quality prediction task. Then the quality score of a patch is obtained by this new feature vector. The framework of our CNN model is demonstrated in Fig. 1. The max pooling layer is introduced to decrease the training parameter size, which also tends to preserve more image texture information. That is consistent with the characteristics of SCIs. Taking distorted SCI patches preprocessed, we feed them into our proposed network to learn their HOG features and quality scores. The specific training objective function is defined as

$$O = \frac{1}{N}\sum_{n=1}^{N}(|{H_p}^n - {H_g}^n|_1 + |{S_p}^n - {S_g}^n|_1) \tag{4}$$

where ${H_p}^n$ represents the predicted HOG features of the n-th image patch in a batch with N patches, ${H_g}^n$ denotes HOG label, ${S_p}^n$ is the predicted score via concatenating the feature vectors from HOG prediction task, and ${S_g}^n$ is the ground truth value. The parameters can be learnt in an end-to-end manner by minimizing the sum of the two L_1-norm loss functions, which comes from the HOG prediction task and quality score prediction task individually.

2.2 Image-Level IQA Score Generation

SCI Patch Selection With the trained multi-task learning CNN model, each input $32 \times 32 \times 3$ SCI patch is predicted for getting a quality score scalar. However, there are a few pure-color patches without containing any contents, whose quality has little impact on the image quality evaluation, even might generate undesired noise for the final quality estimation. Therefore, we need to

get rid of these pure-color patches and select SCI patches containing content information as candidates to obtain the image visual quality score. We mainly refer to [14] to select candidate patches. The result is provided in Fig. 2(d), in which the gray patches belong to the pure-color parts appeared in the original image. Finally, we remove these unimportant patches and reserve these patches that make a difference in terms of the image quality evaluation.

Patch-level Weight Evaluation After selecting suitable candidate SCI patches, there is a need for weighting each patch on account of their different characteristics. As there are two main types of regions in SCI, i.e., textual region and pictorial region. Since HVS is sensitive to texture information, the textual region would always draw more attention than the pictorial region. Correspondingly, textual image patch's quality always supplies a larger impact on the image quality estimation than pictorial patch's quality. Additionally, in some cases, the pictorial region occupies a large proportion in the image, which determines the pictorial patches also have high weights, even higher than those of the textual patches for IQA. Consequently, a comprehensive weight index is needed by taking into account both textual and pictorial weights, and adaptively determining which patch owns higher quality weight for its source image quality evaluation. Based on this observation, we assume that the quality of a SCI is jointly affected by tow factors: the importance of textual and pictorial regions, and the areas of the two regions, which is detailed as follows.

(a) Local weight based on VLSD and ALE: The local weight of each patch is responsible for its own local properties. Concerning that LSD mainly emphasizes textual region and local entropy feature prefers to highlight pictorial region, which was stated earlier, we compute the variance of LSD (VLSD) and the average value of local entropy (ALE) of each patch to reflect their characteristics. Accordingly, the two features of VLSD and the ALE tend to highlight textual and pictorial patches, which is illustrated in Figs. 3(a) and 3(b). The highlighted regions are marked with red boxes to emphasize the difference between the VLSD map and the ALE map. It can be observed that patches with greater value supply more information to image perception. Therefore, these patches deserve to be assigned higher weights for IQA. Correspondingly, the great VLSD value and ALE value can exactly depict and measure local weight of these patches. The two feature values of VLSD and ALE are calculated as

$$VLSD = \frac{1}{N}\sum_{n=1}^{N}\left(\sigma(i,j) - \bar{\sigma}\right)^2 \tag{5}$$

$$ALE = \frac{1}{N}\sum_{n=1}^{N} E(i,j) \tag{6}$$

$$W_l = VLSD^{\rho_1} \times ALE^{\rho_2} \tag{7}$$

where N=1024, which is the number of pixels in the 32×32 image patch, $\sigma(i,j)$ is computed with Eq. (3) for getting LSD map of the patch, and $\bar{\sigma}$ is the mean

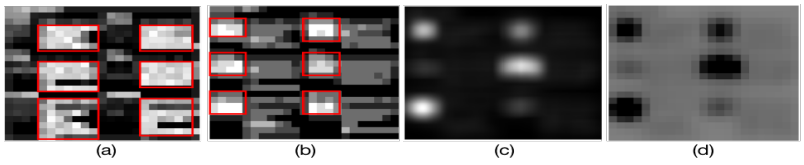

Fig. 3. Examples of local weight and global weight indicators.

value of the LSD map, thus the result of $VLSD$ represents the variance of LSD map of this patch. By the same token, $E(i,j)$ is calculated from the local entropy map of the patch, and then Eq. (6) is applied to compute the average value of the local entropy map. At this point, two weight indicators are obtained for each patch and they need to be combined to form the local weight of this image patch via Eq. (7). The given exponent ρ_1 and ρ_2 are two positive integers that are utilized to adjust the relative importance of $VLSD$ and ALE. In our study, the values of ρ_1 and ρ_2 are variable to adapt to different image databases. We provisionally set these two parameters as 2.0 and 6.0, respectively.

(b) Global weight based on region area: The global weight of each patch aims at the property of the image. According to the type of SCI contents, there are two types of regions: textual region and pictorial region. Facing the fact that humans always pay more attention to the type of the contents covering larger region area instead of the type of contents only occupying a small region area of the image, even if this kind of contents are rich in texture information. For that reason, the content information distribution of the image need to be considered as computing its patches' weights. We call the weight depending on the content region area as global weight, which can be determined through the saliency map. Besides, we observe that the fast saliency map (SM) calculation is a positive way to distinguish the textual regions from the pictorial regions of SCIs, as demonstrated in Fig. 3(c). It's obvious that the saliency map will mostly assign large salient values to pictorial regions instead of textual regions. To put it another way, small salient pixel values in a region suggest that this part is likely to be a textual region. Consequently, we count the number of textual patches and pictorial patches of each SCI respectively, according to the saliency map of SCI. Specifically, we transform the pixel-level salient value to patch-level salient value by computing the maximum value only in each patch for the saliency map. The result is presented in Fig. 3(d) and the calculation is described as

$$S = max\left\{s(i,j)\right\}, i \in [1, N'], j \in [1, N'] \tag{8}$$

where N' is set as the patch size, $s(i,j)$ is saved as pixel-level salient value, and S is saved as patch-level salient value. Then, for each SCI, we count the S values ranging from 0 to 0.03 in order to obtain the number of patches whose contents are inclined to be textual information. Therefore, the remaining patches belong to pictorial patches. Notably, the value of 0.03 is small enough to contain all of textual patches of SCI as much as possible, which is concluded through experiment. Eventually, we obtain the individual amount of pictorial patches and

textual patches in one image and they can represent the global weight of each patch, and we also classify these patches into two categories: textual patches and pictorial patches through judging the S value of each patch. If this patch belongs to a textual patch, the patch's global weight equals to the amount of all textual patches of its source image. Similarly, if this patch is regarded as a pictorial patch, its global weight equals to the amount of all pictorial patches.

Quality Aggregation Given a test distorted screen content image, based on the above analysis, we obtain the quality score of each patch predicted by the multi-task learning network and the corresponding local weight and global weight. A weighted summation method is employed to fuse quality scores of these patches which is calculated as

$$Q = \frac{\sum_{i=1}^{M} Q_i \times (W_{l_i} + W_{g_i})}{\sum_{i=1}^{M} (W_{l_i} + W_{g_i})} \tag{9}$$

where Q_i is the score of the $i-th$ patch and its local weight W_{l_i} is calculated by Eq. (7). Besides, through assessing the type of this patch and count the amount of all patches of its corresponding type in this given image, the global weight W_{g_i} is assigned. M is the number of the patches except pure-color patches of the test image, and Q is the final quality score of the test image.

3 Experimental Results

3.1 Databases and Evaluation Methodology

We employed two widely used SCI databases, i.e., the screen content image quality assessment database (SIQAD) and the screen content image database (SCID), which contain 980 and 1800 high-quality annotated SCIs, respectively. Four measures were leveraged to evaluate the performance of IQA for SCIs: Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Order Correlation Coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC), and Root Mean Square Error (RMSE).

3.2 Performance Comparison

Ablation Studies We conducted ablation experiment on network architecture to examine whether the performance of the proposed multi-task learning model is superior. The baseline model is the quality prediction task without concat layer. Then we added the HOG feature prediction task and concat layer, respectively. The evaluation is illustrated in Fig. 4. The result of the single-task learning model containing quality score prediction task simply is marked in green, named as STL. The multi-task learning model without concat layer is marked in blue, in which the HOG prediction task fails to aid the quality prediction task and only the CNN extractor is shared between the two tasks, named as HOG-MTL, and

the multi-task learning model with concat layer is marked in orange, in which the HOG prediction task successfully aids the quality prediction task, shortly named as HOGAMTL. As presented in Fig. 4, the performance is improved when HOG features are introduced and the accuracy will further increase as the HOG prediction task aids the quality prediction task by the concat layer, which confirms HOG prediction task is effective for visual quality prediction of SCIs.

In order to show the advantage of our proposed quality aggregation algorithm, we conducted another ablation study on weighting strategy. Specifically, each SCI patch is assigned with the same weight (i.e., average weighting strategy), the local weight computed with Eq. (7), the global weight obtained through Eq. (8) and the combination weight via fusing local weight with global weight in Eq. (9), respectively. Table 1 lists the performance evaluation results of different weighting strategies for IQA of SCIs. In this table, the first-ranked, the second-ranked and the third-ranked performance value of each evaluation criteria are boldfaced in red, blue and black, separately. It can be clearly observed from this table, as we assign the combination weight consisting of local weight and global weight to the image patch, the performance will be superior. In contrast, if we get rid of either kind of the two weights, the performance would decrease.

Besides, there is an interesting phenomenon that the local weighting achieves significant improvement comparing with average and global weighting on SIQAD. This could be explained by considering that more massive texts exist in SIQAD than in SCID. If we only adopt global weighting strategy, maybe it will fail to notice the presence of pictorial regions, which leads to the unfairness. By the same logic, if we only adopt average weighting strategy, it will fail to notice the peculiarity of a large number of texts. Consequently, local weighting strategy considering both characteristics of textual and pictorial regions is better.

Overall and individual Performance Comparison In order to verify the overall capability, our proposed NR-IQA model HOGAMTL was compared with multiple FR and NR-SCIQA models, including SIQM [10], ESIM, SVQI [8], SFUW [6], GFM [16], MDQGS [7], SQMS [9], BQMS [11], SIQE [12], OSM [15], NRLT [5], IGM [18], PICNN [2], TFSR [17], QOD [13], and CBIQA [1]. In contrast with the NR-IQA of SCIs models, the first three performance figures of

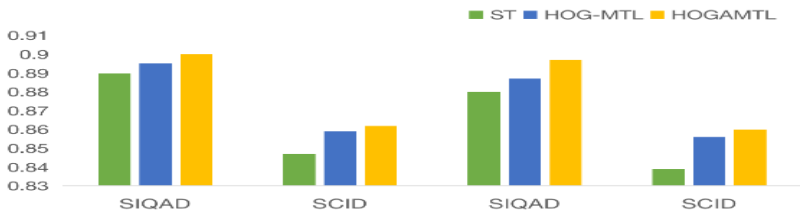

Fig. 4. PLCC and SROCC based on different network models on SIQAD and SCID database. The left two groups denote the PLCC values and the right two groups represent the SROCC values.

Table 1. Evaluation Results Based on Different Weighting Strategy

	SIQAD				SCID			
criteria	SROCC	PLCC	KROCC	RMSE	SROCC	PLCC	KROCC	RMSE
Average	**0.8580**	**0.8648**	**0.6677**	**7.2667**	0.8398	0.8454	0.6444	7.1468
Global	0.8398	0.8499	0.6478	7.6269	**0.8526**	**0.8592**	**0.6602**	**6.8463**
Local	**0.8960**	**0.8984**	**0.7262**	**6.3558**	**0.8542**	**0.8598**	**0.6651**	**6.8332**
Local+Global	**0.8962**	**0.9000**	**0.7265**	**6.3535**	**0.8569**	**0.8613**	**0.6679**	**6.7991**

Table 2. Comparisons among the Proposed and Other FR and NR Models

	Criteria	SIQM FR	ESIM FR	SVQI FR	SFUW FR	GFM FR	MDQGS FR	SQMS FR	BQMS NR	SIQE NR	OSM NR	NRLT NR	IGM NR	PICNN NR	TFSR NR	QOD NR	CBIQA NR	Our
SIQAD	PLCC	0.8520	0.8788	0.8911	0.8910	0.8828	0.8839	0.8870	0.7549	0.7904	0.8306	0.8442	0.8834	0.896	0.8618	**0.9008**	**0.9109**	**0.9000**
	SROCC	0.8450	0.8632	0.8836	0.8800	0.8735	0.8822	0.8803	0.7223	0.7593	0.8007	0.8202	0.8634	**0.897**	0.8354	0.8888	**0.8976**	**0.8962**
	RMSE	7.4936	6.8310	6.4965	6.4990	6.7234	6.6951	6.6110	9.3042	8.7899	7.9331	7.5957	-	6.790	7.4951	**6.2258**	**5.8930**	**6.3535**
SCID	PLCC	0.8303	**0.8630**	0.8604	0.8590	**0.8760**	-	0.8557	0.6487	0.6371	-	0.6625	-	-	**0.8017**	-	**0.8531**	**0.8613**
	SROCC	0.8086	0.8478	0.8386	**0.8950**	**0.8759**	-	0.8320	0.6138	0.6034	-	0.6454	-	-	**0.7840**	-	**0.8377**	**0.8569**
	RMSE	7.8920	7.1552	7.2178	7.3100	6.8310	-	7.3276	10.7787	10.9202	-	10.6452	-	-	**8.8041**	-	**7.3930**	**6.7991**
Direct	PLCC	0.8412	0.8709	0.8758	0.875	0.8794	-	0.8714	0.7018	0.7138	-	0.7534	-	-	**0.8318**	-	**0.882**	**0.8807**
Average	SROCC	0.8268	0.8555	0.8611	0.8875	0.8747	-	0.8562	0.6681	0.6814	-	0.7328	-	-	**0.8097**	-	**0.8677**	**0.8766**
Weighted	PLCC	0.8379	0.8686	0.8712	0.8703	0.8783	-	0.8667	0.6861	0.6911	-	0.7266	-	-	**0.8229**	-	**0.8735**	**0.875**
Average	SROCC	0.8214	0.8532	0.8545	0.8897	0.8751	-	0.849	0.652	0.6584	-	0.707	-	-	**0.8021**	-	**0.8588**	**0.8708**

Table 3. Performance Comparisons on Seven Distortion Types on SIQAD

	Distortions	SIQM FR	ESIM FR	SVQI FR	SFUW FR	GFM FR	MDQGS FR	SQMS FR	BQMS NR	SIQE NR	NRLT NR	TFSR NR	QOD NR	CBIQA NR	Our
PLCC	GN	0.8921	0.8891	0.9031	0.8870	0.8990	0.8982	0.900	0.8372	0.0883	0.9131	**0.9291**	0.913	**0.9317**	**0.9177**
	GB	0.9124	0.9234	0.9132	0.9230	0.9143	0.9195	0.912	0.7558	0.8033	0.8949	**0.9367**	**0.925**	0.9148	**0.9535**
	MB	0.8565	0.8886	0.8722	0.8780	0.8662	0.8421	0.867	0.7237	0.7810	**0.8993**	**0.9243**	0.889	0.8846	**0.9282**
	CC	0.7902	0.7641	0.8087	0.8290	0.8107	0.8011	0.803	0.7209	0.6030	**0.8131**	0.6563	**0.837**	**0.9229**	0.7049
	JC	0.7717	0.7999	0.7953	0.7570	0.8398	0.7885	0.786	0.7653	**0.8339**	0.7932	0.8334	0.830	**0.9036**	**0.9138**
	J2C	0.7940	0.7888	0.8342	0.8150	0.8486	0.8606	0.826	0.7909	**0.8535**	0.6848	0.8347	0.818	**0.9143**	**0.9368**
	LSC	0.7204	0.7915	0.8283	0.7590	0.8288	0.8316	0.813	0.8427	**0.8921**	0.7228	0.8069	**0.867**	**0.9294**	0.8128
	ALL	0.8520	0.8788	0.8911	0.8910	0.8828	0.8839	0.8870	0.7549	0.7904	0.8842	0.8616	**0.9008**	**0.9109**	**0.9000**
SROCC	GN	0.8711	0.8757	0.8909	0.8690	0.8795	0.8882	0.886	0.8346	0.8280	0.8966	**0.9144**	**0.905**	**0.9143**	0.8692
	GB	0.9102	0.9239	0.9129	0.9170	0.9132	0.9192	0.915	0.7627	0.7942	0.8812	**0.9311**	**0.916**	0.8971	**0.9365**
	MB	0.8401	0.8938	0.8753	0.8740	0.8699	0.8345	0.869	0.7176	0.7748	**0.8919**	**0.9148**	0.871	0.8708	**0.9184**
	CC	0.7055	0.6108	0.7131	0.7220	0.7038	0.6644	0.695	**0.7260**	**0.8199**	0.7072	0.6498	0.700	**0.9075**	0.5835
	JC	0.7754	0.7989	0.7925	0.7500	0.8434	0.7856	0.789	0.7661	**0.8388**	0.7698	0.8377	0.815	**0.8848**	**0.8922**
	J2C	0.7771	0.7827	0.8282	0.8120	0.8444	0.8622	0.819	0.7919	**0.8493**	0.6761	0.8354	0.795	**0.8911**	**0.9562**
	LSC	0.7255	0.7958	0.8412	0.7540	0.8445	0.8513	0.829	0.8267	**0.8843**	0.6978	0.7948	**0.882**	**0.9046**	0.7898
	ALL	0.8450	0.8632	0.8836	0.8800	0.8735	0.8822	0.8803	0.7223	0.7593	0.8202	0.8354	**0.8888**	**0.8976**	**0.8962**
RMSE	GN	7.0165	6.8272	6.4044	6.8760	6.6835	6.5576	6.921	8.1615	19.0113	-	**5.3105**	6.150	**5.3292**	**5.9159**
	GB	5.8367	5.8270	6.1550	5.5920	6.1459	5.9639	6.611	8.8390	8.2689	-	**5.2141**	5.772	**5.3767**	**4.4336**
	MB	6.0869	5.9639	6.3604	6.2360	6.5184	7.0121	7.204	9.2398	8.6522	-	**5.5266**	**5.762**	6.0794	**4.6997**
	CC	8.1079	8.1141	7.3996	7.0480	7.3638	7.5284	7.743	**9.2114**	12.4155	-	10.5005	**6.939**	**5.0375**	9.9351
	JC	5.6548	5.6401	5.6969	6.1430	5.1009	5.7787	5.983	8.5874	7.3633	-	**5.2541**	**5.460**	5.5912	**3.7027**
	J2C	6.0820	6.3877	5.7309	6.0230	5.4985	5.2930	6.050	8.4164	7.1150	-	**5.6377**	6.000	**5.4480**	**3.1189**
	LSC	5.3576	5.2150	47751	5.5550	4.7736	4.7382	5.104	7.8336	6.5744	-	5.6217	**4.338**	**5.2539**	**5.3892**
	ALL	7.4936	6.8310	6.4965	6.4990	6.7234	6.6951	6.6110	9.3042	8.7899	0.8202	7.4951	**6.2258**	**5.8930**	**6.3535**
	Number	-	-	-	-	-	-	-	0	0	0	2	1	9	**12**

each measurement criterion (i.e., PLCC, SROCC and RMSE) in each row are indicated in bold, and from the first to the third are individually marked red, blue and black. It can be obviously observed that the proposed HOGAMTL yields the best overall performance on the SCID database, compared with other state-of-the-art NR-SCIQA methods. On the SIQAD database, our method achieves the third-place overall performance but almost comparable to the top two models. Certainly, HOGAMTL even surpasses various FR-IQA of SCIs models. Specifically, the performance of HOGAMTL is higher than all advanced FR-IQA of

Table 4. Performance Comparisons on Nine Distortion Types on SCID

	Distortions	SIQM (FR)	ESIM (FR)	SVQI (FR)	GFM (FR)	SQMS (FR)	CBIQA (NR)	Our (NR)
PLCC	GN	0.9269	**0.9563**	0.9362	**0.9497**	0.9298	0.9182	**0.9722**
	GB	**0.9266**	0.8700	0.9130	**0.9156**	0.9081	**0.9296**	0.8579
	MB	**0.9152**	0.8824	0.8997	**0.9023**	0.8968	**0.903**	0.8983
	CC	0.7821	0.7908	0.8266	**0.8787**	**0.8441**	**0.9067**	0.7584
	JC	0.9226	**0.9421**	0.9356	**0.9392**	0.9302	0.8992	**0.9558**
	J2C	0.9076	**0.9457**	**0.9513**	0.9226	**0.9468**	0.9028	0.9410
	CSC	0.0683	0.0694	0.0919	**0.8728**	0.0628	**0.8527**	**0.6167**
	CQD	0.8385	**0.9005**	**0.9047**	0.8928	0.8986	**0.9042**	0.7842
	HEVC	0.8316	**0.9108**	0.8496	**0.8740**	**0.8515**	0.8363	0.8379
	ALL	0.8303	**0.8630**	0.8604	**0.8760**	0.8557	0.8531	**0.8613**
SROCC	GN	0.9133	**0.9460**	0.9191	**0.9370**	0.9155	0.8962	**0.9641**
	GB	**0.9232**	0.8699	0.9079	**0.9081**	**0.9079**	0.9063	0.8401
	MB	**0.9006**	0.8608	**0.8842**	0.8892	0.8814	0.8725	**0.9107**
	CC	0.7435	0.6182	0.7705	**0.8225**	**0.8027**	**0.8797**	0.5300
	JC	0.9158	**0.9455**	**0.9287**	0.9281	0.9236	0.8959	**0.9558**
	J2C	0.8935	**0.9359**	**0.9367**	0.9085	**0.9320**	0.8923	0.9006
	CSC	0.0617	0.1037	0.0790	**0.8736**	0.0814	**0.8396**	**0.5350**
	CQD	0.8301	0.8868	**0.8957**	0.8907	**0.8913**	**0.8955**	0.7505
	HEVC	0.8517	**0.9036**	0.8665	**0.8712**	**0.8667**	0.8261	0.8569
	ALL	0.8086	**0.8478**	0.8386	**0.8759**	0.8320	0.8377	**0.8569**
RMSE	GN	4.8222	**3.6760**	4.4179	**3.9378**	4.6250	4.9014	**2.7681**
	GB	**4.0989**	5.2213	4.3194	**4.2566**	4.4336	**4.3107**	5.3026
	MB	**4.7388**	5.1431	4.7709	**4.6121**	4.8352	5.1268	**4.6197**
	CC	6.1281	5.4790	**5.0374**	**4.2732**	**4.7995**	5.0742	5.3783
	JC	6.7341	**5.0373**	5.3055	5.2011	5.5181	**4.5375**	**4.4987**
	J2C	7.2951	5.1695	**4.9058**	6.1385	5.1191	**4.2713**	**4.8037**
	CSC	9.8394	9.8156	9.7977	**4.8031**	9.8199	**5.1742**	**7.1440**
	CQD	7.1976	**5.5607**	**5.4481**	5.7592	5.6110	**4.9035**	8.7329
	HEVC	8.197	**5.7446**	7.3381	6.7590	7.2938	**5.3346**	**6.2844**
	ALL	7.8920	**7.1552**	7.2178	**6.8310**	7.3276	7.3930	**6.7991**
	Number	3	2	4	7	0	6	**8**

SCIs model listed in Table 2 on the SIQAD database in three metrics, and it also beats nearly all FR-IQA of SCIs models on the SCID database except ESIM, SFUW and GFM highlighted in green. For the sake of comprehensive performance comparisons over multiple databases, table 2 exhibits the average performances of different IQA methods on these two databases. It can be obviously observed that the proposed HOGAMTL almost yields the highest PLCC and SROCC in both Direct Average and Weighted Average performance comparisons apart from PLCC value of Direct Average, yet the metric value (0.8807) is intensely approximate to the top value (0.882). Consequently, it's obvious that HOGAMTL offers comprehensive performance over multiple databases.

To more comprehensively evaluate each IQA model's ability on assessing image quality's degradations caused by each distortion type, Table 3 reports the results of comparison experiment conducted on the SIQAD database. Compared with these FR-IQA of SCIs models, HOGAMTL surpasses most of them excluding the performance on CC distortion type. Among these NR-IQA of SCIs models, the proposed HOGAMTL yields the most top-three performances (18 times) same as model CBIQA. Yet HOGAMTL is among the first-place models

Table 5. Performance in Cross Database Validation

	Training (SIQAD)	
Testing (SCID)	PLCC	SROCC
GN	0.8434	0.8084
GB	0.7637	0.7342
MB	0.8865	0.8745
CC	0.7517	0.6627
JC	0.8494	0.8258
J2C	0.8621	0.8445
ALL	0.8267	0.8186

12 times beyond CBIQA (9 times) with 3 times, which is presented in the last row of Table 3 where "Number" implies the number of occurrences as first-ranked model. That proves the superiority capability of our proposed method. Table 4 exhibits the results of different distorted types for the HOGAMTL and compared methods on SCID. Differently from before, this time the top three models are evaluated under the participation of FR-IQA of SCIs models. As shown in Table 4, the proposed model delivers excellent performance with 8 times among the first-place models and renders more promising performance on SCID database compared with results with method CBIQA on SIQAD, which further verifies the robustness of HOGAMTL. We also trained the multi-task learning network model on the whole SIQAD database and tested on SCID database. As shown in Table 5, there is a acceptable performance decline compared to training on SCID database, which represents a strong generalization performance of our model.

4 Conclusion

In this paper, a multi-task deep learning model had been proposed for NR-SCIQA. We first introduced the HOG prediction task to our multi-task learning model and let this task to aid the quality prediction task, then we designed a quality aggregation algorithm aiming at fusing local quality scores of image patches to obtain the final quality score of screen content image. The comparison experiments verified the superior performance and generalization of our method.

Albeit our method has achieved outstanding performance, the performance of CC distorted SCI quality prediction still has a relatively large room for improvement. For the future work, we will continue to explore better texture features, even contrast features to aid the deep learning more precisely.

References

1. Bai, Y., Yu, M., Jiang, Q., Jiang, G., Zhu, Z.: Learning content-specific codebooks for blind quality assessment of screen content images. Signal Processing **161**(20), 248–258 (2019)
2. Chen, J., Shen, L., Zheng, L., Jiang, X.: Naturalization module in neural networks for screen content image quality assessment. IEEE Signal Processing Letters **25**(11), 1685–1689 (2018)

3. Cheng, Z., Takeuchi, M., Kanai, K., Katto, J.: A fast no-reference screen content image quality prediction using convolutional neural networks. In: Proceedings of the IEEE International Conference on Multimedia & Expo Workshops. pp. 1–6 (2018)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 886–893 (2005)
5. Fang, Y., Yan, J., Li, L., Wu, J., Lin, W.: No reference quality assessment for screen content images with both local and global feature representation. IEEE Transactions on Image Processing **27**(4), 1600–1610 (2017)
6. Fang, Y., Yan, J., Liu, J., Wang, S., Li, Q., Guo, Z.: Objective quality assessment of screen content images by uncertainty weighting. IEEE Transactions on Image Processing **26**(4), 2016–2027 (2017)
7. Fu, Y., Zeng, H., Ma, L., Ni, Z., Zhu, J., Ma, K.K.: Screen content image quality assessment using multi-scale difference of gaussian. IEEE Transactions on Circuits and Systems for Video Technology **28**(9), 2428–2432 (2018)
8. Gu, K., Qiao, J., Min, X., Yue, G., Lin, W., Thalmann, D.: Evaluating quality of screen content images via structural variation analysis. IEEE Transactions on Visualization and Computer Graphics **24**(10), 2689–2701 (2017)
9. Gu, K., Wang, S., Yang, H., Lin, W., Zhai, G., Yang, X., Zhang, W.: Saliency-guided quality assessment of screen content images. IEEE Transactions on Multimedia **18**(6), 1098–1110 (2016)
10. Gu, K., Wang, S., Zhai, G., Ma, S., Lin, W.: Screen image quality assessment incorporating structural degradation measurement. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS). pp. 125–128 (2015)
11. Gu, K., Zhai, G., Lin, W., Yang, X., Zhang, W.: Learning a blind quality evaluation engine of screen content images. Neurocomputing **196**, 140–149 (2016)
12. Gu, K., Zhou, J., Qiao, J.F., Zhai, G., Lin, W., Bovik, A.C.: No-reference quality assessment of screen content pictures. IEEE Transactions on Image Processing **26**(8), 4005–4018 (2017)
13. Jiang, X., Shen, L., Feng, G., Yu, L., An, P.: Deep optimization model for screen content image quality assessment using neural networks. arXiv preprint arXiv:1903.00705 (2019)
14. Kim, J., Lee, S.: Deep learning of human visual sensitivity in image quality assessment framework. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1676–1684 (2017)
15. Lu, N., Li, G.: Blind quality assessment for screen content images by orientation selectivity mechanism. Signal Processing **145**, 225–232 (2018)
16. Ni, Z., Zeng, H., Ma, L., Hou, J., Chen, J., Ma, K.K.: A gabor feature-based quality assessment model for the screen content images. IEEE Transactions on Image Processing **27**(9), 4516–4528 (2018)
17. Yang, J., Liu, J., Jiang, B., Lu, W.: No reference quality evaluation for screen content images considering texture feature based on sparse representation. Signal Processing **153**(89), 336–347 (2018)
18. Yue, G., Hou, C., Yan, W., Choi, L.K., Zhou, T., Hou, Y.: Blind quality assessment for screen content images via convolutional neural network. Digital Signal Processing **91**, 21–30 (2019)
19. Zuo, L., Wang, H., Fu, J.: Screen content image quality assessment via convolutional neural network. In: Proceedings of the IEEE International Conference on Image Processing. pp. 2082–2086 (2016)