No-reference image quality assessment with region ofi nterest

Chenge Liu, Shanzhi Zhang(chenge@ualberta.ca,shanzhi1@ualberta.ca)

Abstract

In recent years, the most of No-reference image quality assessment algorithms concentrate on the assessment of the unified image which uses a huge amount of data set to train the neural networks and receive any type of image as input to give the score of this image.

This kind of algorithm gives a satisfying result of overall image quality assessment. However, for the specific type image, we assume those algorithm will perform worse. [1] For example, in a car image, people will concentrate on the car rather than the background of this image, if the quality of the background is bad, those algorithms will give a lower score than people give. [2, 3] We are trying to extract the region of interest at the first, then use no-reference image quality assessment with a deep learning algorithm to give the score of this specific region. Then using a weight of this score to improve the original score of the whole image and make it closer to the score given by real human expert.

Introduction

In the past few decades, the quality assessment of image has wide practicability in many fields, such as image compression, video code, video monitoring. There is an increasing demand for highly efficient and reliable image quality assessment algorithm.

IQA can be divided into the subjective evaluation and objective evaluation. Subjective evaluation is to evaluate the image quality of people's subjective perception. First, the original reference image and the distorted image are given, and the annotator is asked to Score the distorted image. Objective evaluation using mathematical model to give evaluation value. However, the subjective assessment is time-consuming and laborious, which is not feasible in practical application, and the subjective experiment is affected by many factors such as viewing distance, display equipment, lighting conditions, the observer's visual ability and mood. Therefore, it is necessary to design a mathematical model that can predict subjective mass automatically and accurately.

IQA(image quality assessment) is generally divided into three categories according to the amount of information provided by the original Reference image: Full Reference-IQA, Reduced Reference-IQA, and No Reference-IQA. The main differences between those three methods shown in Fig. 1. Full reference has both original (without distortion) images and distorted images, with low difficulty. The core is to compare the information content or feature similarity of the two images, which is a mature research direction.No reference only has distorted images, which is of high difficulty. It is a research hotspot in recent years, and also the most challenging problem in IQA. Reduced reference has only partial information of the original image or partial features extracted from the reference image. This method is between full reference and no reference, and any full reference and no reference methods can be converted into reduced reference methods after proper processing. Furthermore, no reference algorithms can be subdivided into two categories: one that studies specific types of image quality, such as estimation of blur, block effect, and noise severity, and the other that estimates non-specific types of image quality. In general, reference images cannot be provided in practical applications, so no reference has the most practical value, has a wide range of applications, and is very convenient to use. At the same time, it also becomes a difficult research object due to the ever-changing image content and no reference.

In recent years, deep learning has attracted the attention of researchers and achieved great success in various computer vision tasks. Specifically, CNN(convolutional neural network) has demonstrated superior performance on many standard object recognition benchmarks. One of the advantages of CNN is that it can directly take the original image as input and integrate feature learning into the training process. CNN has a deep structure that can effectively learn complex mappings while requiring minimal domain knowledge. And it shows high performance in the no-reference quality assessment area. And some new deep learning-based no reference algorithm even performs better than some full reference algorithm. [4]

And in this paper, we are trying to combine the method of region of interest and deep learning network. Dose region of interest helps deep learning network give a more accurate score? It should be an interesting topic to research. Actually, some people had already done some work in this area a few years ago, it showed some improvements of combination of region of interest and the assessment algorithm. [5] However, People did not use deep learning as an image quality assessment method at that point. In recent years, deep learning IQA becomes the most particle and accurate method, so we want to see if we can have a better result with those two methods. Our final target is combining region of interest and deep learning image quality assessment to see what improvements will show.

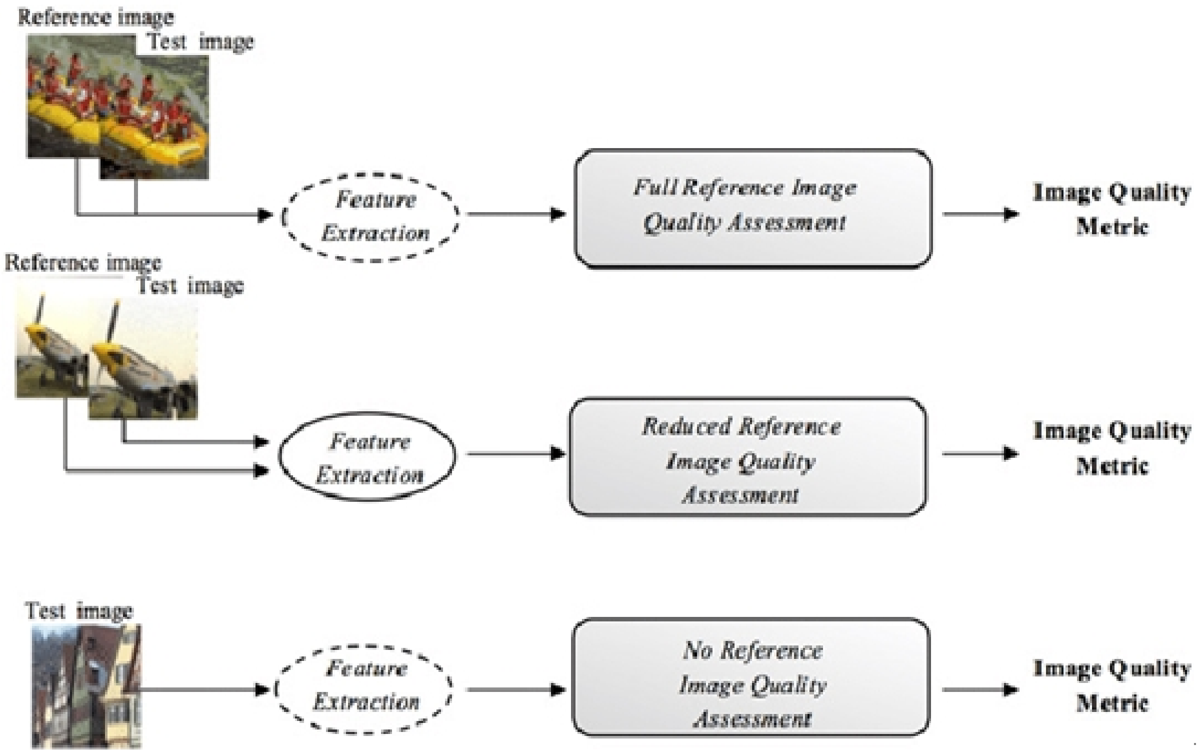

Figure 1: Relationship between subjective and objective image quality assessment.

literature review

Our literature review has two parts. The first part is summarizing the existing algorithm about choosing a region of interest in a given image. The second part is about image and video quality assessment algorithms. During our research, we did not see any existing algorithm combine the two neural networks to extract ROI and image quality assessment. Therefore, we want to see what will happen if we combine a visual saliency map prediction neural network with a deep learning neural network for image quality assessment. We accept this method will have more accuracy and good average performance on different data sets.

Review of visual saliency map prediction algorithm

During our research, we discovered many methods for extracting ROI and image segmentation. All those algorithms have a common target which is finding the region that attracts the most human attention. And we can use a Visual saliency map to represent how many attended the person will pay to each area in the image. Therefore, how to predict visual saliency maps accurately and effectively is the most important part to solve in the extracting ROI problem. [6]

To solve this problem, at the beginning we find people trying to create a model Jonathan Harel, Christof Koch and Pietro Perona performed a three-stage model to predict visual saliency for a given image. [7] First, they extract features in the given graph by some linear filter. Then use this result map as input for the next stage to calculate the activation map. This map shows the difference between two adjacent pixels. Then they use a new formula called Markovian Approach. This approach predicts visual saliency by calculating the dissimilarity of every pixel with its surrounding pixels. And they called this method "organic" because every pixel in the image is like the neurons in the human body. If the pixel is different from the input numbers from the previous pixel, it will have a large value to show its visual saliency. And this method is the basement of many algorithms nowadays, but because it only uses one formula to express the visual saliency difference in the image, it does not perform well as real human eyes result.

owever, for those bottom-up calculation algorithms, the output usually does not match very well with the result by eye-tracking of real humans. The reason is those algorithms usually need many design parameters, and people can not realize which parameter is more important easily. [8]

Then in terms of the high development of supervised learning, people try to use new technology to make predictions in visual saliency maps. T. Judd, K. Ehinger, F. Durand and A. Torralba give a large database of eye-tracking experiments with labels and analysis and use supervised learning to train a model for predicting visual saliency maps. [8] In their work, they use three levels of features as the input of their network. First, they use Intensity, orientation and colour contrast as the lower level features, those three features are commonly used in many algorithms about predicting visual saliency. And then they use a horizon line detector as middle-level features, this is because all objects are laying on the surface of the earth, this is an interesting point and as result, these features work well during the training network. In the end, they used a face detector as a high-level feature. As a result, this supervised learning network performed well in the human imagination and it can achieve about 90 percent accuracy in a specific type of image. However, for the average performance, it does not have an ideal performance because the supervised learning is only concentrated on one specific type data set, it will cause an overfitting problem which makes this network do worse in other type input data.

In recent years, the deep learning and neural network has had high improvements, people find its good probability to find the parameters and relationship in the input image and output ROI. Therefore many algorithms based on deep learning have occurred. And those algorithms show a very good performance in predicting visual saliency maps and extract ROI of the given image. As our research, we found two typical deep learning algorithms in this area.

The first one is in 2014, Eleonora Vig, Michael Dorr and David Cox came up with a neural network method to predict visual saliency map. And this network will calculate the parameters automatically rather than need parameters designed by experts. It performed really well in the MIT300 data set and beyond all the other algorithms at that time. [9] They use a set of models as a network and trained them with three-level features. Then use linear SVM to get the final output. Actually it is not a totally deep learning method because they also use many supervised learning during the training. However, it still shows the good potential ability to deal with this type of problem. And in this research, they tried to label every pixel's saliency in the input image, it shows more accuracy as a result. Because they did not use a really efficient training algorithm, their network needed lots of data and time to train. We will try to train our network with some recent algorithms and some training skills in our implementation to improve this part.

After four years later, there is a new algorithm with a convolution neural network called SalGAN. [10] In this method, they used one generator and one discriminator which are two neural networks. In the generator, they used max-pooling layers to reduce the size of input then using upsampling layers followed by convolutional filters to construct output. In the discriminator part, the network is composed of six kernel convolutions interspersed with three pooling layers and followed by three fully connected layers. Then they trained over a perceptual loss [11] which combines content and adversarial loss. The content loss is computed on every pixel by comparing the predicted saliency map with its ground truth. It will output the difference between the prediction and real value. And for the adversarial loss part, they used a former network called Generative adversarial networks (GANs) [12]. They used this network to generate the loss of the prediction of the discriminator over the generated saliency map. And this is the first network that was trained by adversarial samples at that time. It shows that adversarial training can improve a deep neural network remarkably. As a result, those two types of loss used to train the model, give the network a very good probability to find the accurate visual saliency map and improve the efficiency of the training process. Then they used 15000 images(a batch size of 32) to train this network. This neural network performs really well in many different data sets and it is easy to train and set. So we decide to use this neural network for our extracting ROI part, we will use this network to predict visual saliency maps then segment that area has more visual saliency. Use this segmentation image as the input of our next step which is image quality assessment. Therefore, we will try to use this tool for detecting ROI automatically:

SalGAN: Visual Saliency Prediction with Adversarial Networks
(`https://imatge-upc.github.io/salgan/`)

SalGAN is a deep convolutional neural network for visual saliency map prediction which trained with adversarial data. It shows a good performance on many famous data set (eg.MIT300). And it is an open-source neural network, it also contains training data and whole parameter sets on their website.

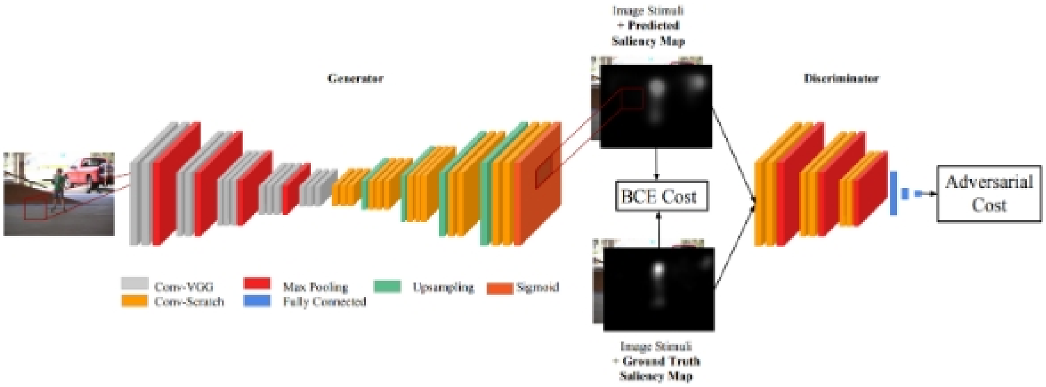

Figure 2: The architecture of the deep neural network used in SalGAN, and the process of training network.

Reviews for image quality assessment algorithm:

During our research,we found that according to how much reference image information is needed for image visual quality calculation, objective quality evaluation algorithms are divided into three categories.The full reference quality evaluation algorithm [13], the semi-reference quality evaluation algorithm [14], and the non-reference quality evaluation algorithm [15]. The full reference and semi-reference quality evaluation algorithms mainly quantify the difference between the reference image and the distorted image by analyzing the visual features of the image to calculate the visual quality of the distorted image. Compared with the full reference quality evaluation algorithm, the non-reference quality evaluation algorithm does not need any information of the reference image when calculating the visual quality of the distorted image, which has been a current study area.

0.1 No-reference image quality evaluation algorithm based on supervised learning

The non-reference image quality evaluation algorithm based on supervised learning mainly includes machine learning and deep learning methods. The method based on traditional machine learning aims to design effective visual feature expression methods. While the method based on deep learning mainly use neural network, learns visual features of images to construct image quality evaluation model, or directly learns function expression of distorted images to image visual quality through end-to-end.

Based on the traditional machine learning method: In 2012, Mittal et al. Proposed to use the natural scene statistics (NSS) features of the image in the airspace to calculate the visual quality of the image. The method considers that the image with good visual quality has a certain brightness value. This kind of distribution, and the presence of noise will destroy the distribution of the brightness value. The author uses the parameters of the fitted Gaussian model as the features of the distorted image, and uses support vector regression to train the mapping model between the features and the visual quality of the image. [15]. Different from the above method, Fang et al. Research found that the moment features (mean, variance, kurtosis, and skewness) and information entropy of the image showed a certain distribution, and proposed to use the output value of the distribution function obtained by the fitting as the characteristics of the distorted image To build a reference-free image quality evaluation model [16]. Both the methods proposed by Mittal et al. And Fang et al. Need to fit a function distribution to extract the relevant features of the image. Different from the above two methods, Li et al. Proposed the use of histograms to represent the statistical characteristics of brightness and structural statistics. The authors transformed the characteristic maps into feature vectors through histograms to build a reference-free image quality evaluation model [17]. In [18], Fang et al. Constructed a screen image visual quality evaluation model based on the characteristics of the screen image (including the image part and the text part), fused the direction information, and extracted the statistical characteristics of brightness and high-order derivative texture features through histograms.

A method based on deep learning: In 2014, Kang et al. Proposed to use a convolutional neural network (CNN) to build an image quality evaluation model, which only includes two convolutional layers and three fully connected layers [19]. Hou et al. Proposed the use of a fully connected neural network to build an image quality evaluation model. The author took the NSS features of the image as input, trained a deep classification network, and divided the visual quality of the image into five categories, including good, good, average, poor, and worst.And the category label of the image is converted into the visual quality of the image through a weighting strategy [20]. Bosse et al. Proposed the use of a deep convolutional neural network VGG16 [21] to build an image quality evaluation model [22]. The model input is consistent with VGG16, and the weight prediction value and quality score prediction value of each image block are fused to obtain the image quality. fraction. Different from the above-mentioned single task end-to-end deep learning model, in [23], Ma et al. Proposed an end-to-end deep convolutional neural network model (MEON) based on multi-task optimization. The idea of this work s derived from BIQI [24], which incorporates image distortion information during the training phase.In MEON, the first stage is to train a classification network of image distortion type; the second stage is a multi-task learning network. The convolution layer framework and the classification network are consistent. The parameters of the classification network are used to initialize the parameters of the second stage training task. After the convolutional layer is a two-way fully connected network, the probability and corresponding quality score of the image belonging to a certain type of distortion are calculated respectively. Finally, the two feature vectors are fused to calculate the visual quality of the image. In [4], Liu et al. Proposed to use ranking learning to obtain the initial parameters of the deep learning network. Based on the ranking learning, fine-tune and get the final image quality model.

0.2 No referenced image quality evaluation algorithm based on unsupervised learning

Similar to supervised learning, no referenced image quality evaluation algorithms based on unsupervised learning mainly include methods based on traditional machine learning and deep learning.

A method based on traditional machine learning: In 2013, Mittal et al. Proposed an unsupervised method based on their previous work to achieve the non-reference image quality evaluation NIQE [25]. A multivariate Gaussian (MVG) model is obtained by fitting the NSS features of a good quality image block. Based on NIQE, Zhang et al. Proposed an Integrated Local NIQE (IL-NIQE) algorithm by integrating structural statistical features, multi-scale direction and frequency statistics, and color statistical features [26]. In 2013, Xue et al. Proposed an image quality evaluation method based on image-perceived mass perception clustering [27]. The quality score of each image block is calculated through the full reference image quality evaluation, and divided into L categories according to the range of quality scores of all image blocks. Based on the extracted structural features, the K-means clustering algorithm The image blocks are divided into K sub-categories. Among them, each small class corresponds to a cluster center, which is an image block (structural features and visual quality score). Given a distorted image, extract the image block, and calculate the distance between the image block and the clustering center of each small class in the L large class. The visual scores of the smallest cluster centers in each category are combined to calculate the scores of the image blocks. Finally, using the average weighting strategy, the visual quality score of the distorted image is obtained. In [28], Zhang et al. First classified the distortion types of the image, then extracted the NSS features for different types of distortion, trained them using SVR to get the corresponding regression model, estimated the distortion parameters and saved. Finally, the visual quality score of the distorted image is calculated.

Method based on deep learning: Ma et al. Designed a fully connected neural network image quality evaluation model based on weight sharing [29]. The input is two images, and the visual characteristics of the images are extracted using the method CORNIA [30]. The image quality evaluation model is trained by blending the uncertainty of the visual differences of the images and the quality of the visual quality of the images. In , Lin et al. Proposed an image quality evaluation method based on a generative adversarial network (GAN). The method includes generating a network and discriminating a network. The goal of generating the network is to generate a false reference image, and the goal of discriminating the network is to distinguish the generated false reference image from a true reference image.

0.3 No referenced image quality evaluation algorithm based on deep learning neural network

In 2017, Xialei Liu, Joost van de Weijer and Andrew D. Bagdanov came up with a new method for no-reference image quality assessment, which is a neural network trained by rank images based on image quality by using synthetically generated distortions for which image quality is known.[8] They provided a new way to use the data set that is not large enough, and this deep learning network trained by this new method performed really well on many data sets, it is even better than some existing full-reference image quality assessment method. First, they apply different level Gaussian blur (or any other distortion) to the reference image in the data set, and get a set which includes those reference images with different level distortions. Then they can easily get the rank of this image set according to the level of distortion applied to the image. After that, they use those ranked sets to train their network. They use Siamese network to learn from image rankings, which is a network with two identical network branches and a loss module. However, this network has a backward which is a complex computation, it will cost lots of time to get a well-performed model. According to this problem, they come up with an Efficient Siamese backpropagation. They reduce the number of times image pass through the network and consider all possible pairs in the loss computation layer. As a result, they speed up the process of convergence without losing accuracy. After the first training, they use a single branch from the network for fine-tuning. This branch trained by the image already has the score that people give with the predicted score from the network. As a result, they achieved at least 5 percent improvements in the best no-reference image quality assessment method at that time.

However, they did not mention the ROI effect on the image quality assessment. They do not consider the visual saliency as an important aspect of the assessment process. So we will try to use one no-reference image quality assessment method combined with the ROI extract method which was mentioned before to try to get a better performance.

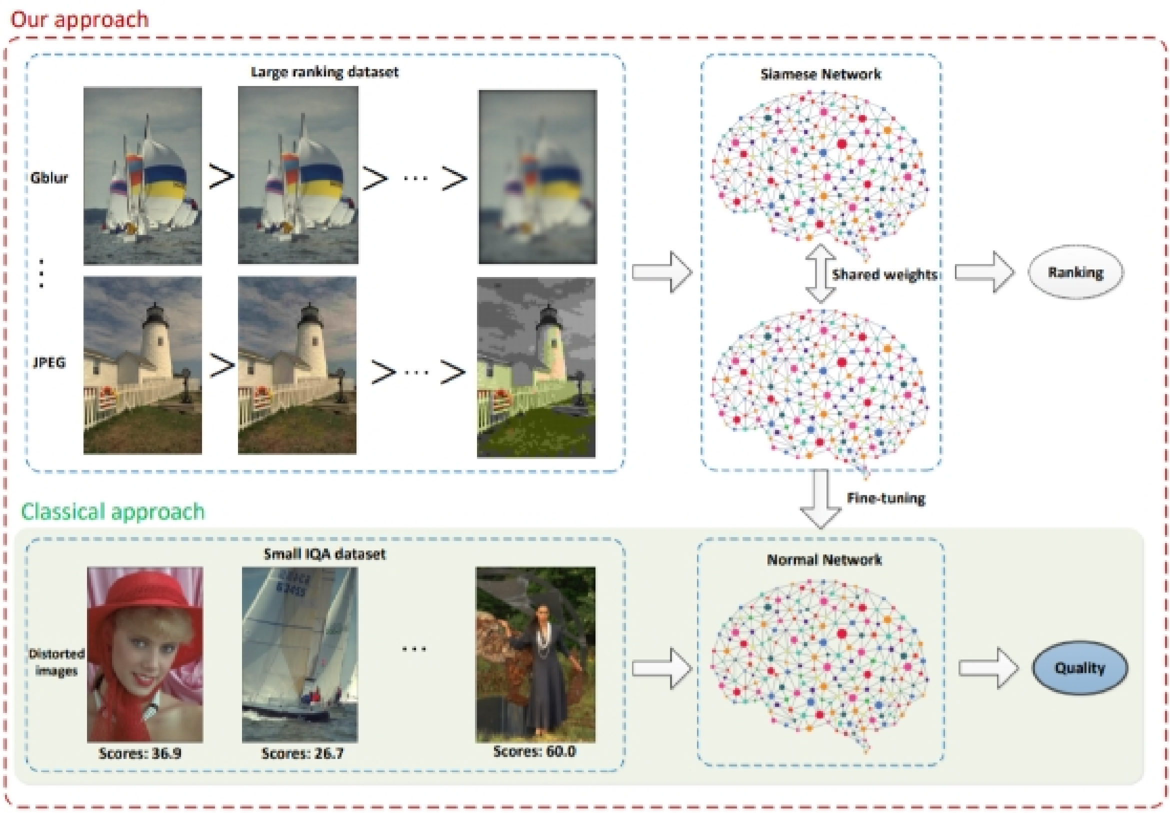

Figure 3: The difference between the classical approach trains a deep CNN with the IQA score and RankIQA

RankIQA: Learning from Rankings for No-reference Image Quality Assessment (https://github.com/xialeiliu/RankIQA)

RankIQA is a no-reference image quality assessment deep learning network trained by image rankings. This network has a good performance on the small size training data and it can achieve a really good performance in many famous image quality assessment data sets. They have per trained network on their website.

Brief Summary of Existing Work

In this section, we will briefly introduce the existing work of video quality assessment and separate them into 3 parts. The indicators of video assessment will be shown in Section 0.4. Section 0.5 shows the traditional algorithm that used in video assessment. Recent work such as deep learning will represent in Section 0.6.

0.4 Indicators of video assessment

There are many indicators for measuring the image quality, and for each indicator has its own characteristics. Usually, we use the differences and correlations between the objective values of the model and the subjective values of the observation objects to get IQA. Two common evaluation indicators are Linear Correlation Coefficient (LCC) and Spearman's Rank Order Correlation Coefficient (SROCC). LCC describes the linear correlation between subjective and objective assessments and is defined as:

$$LCC = \frac{\sum_{i=1}^{N}(y_i-\bar{y})(\overline{\widehat{y_i}}-\bar{\bar{y}})}{\sqrt[2]{\sum_{i=1}^{N}(y_i-\bar{y})^2}\sqrt[2]{\sum_{i=1}^{N}(\overline{\widehat{y_i}}-\bar{\bar{y}}}} \quad (1)$$

SROCC measure the monotonicity of algorithm predictions:

$$SROCC = 1 - \frac{6\sum_{i=1}^{N}(v_i - p_i)^2}{N(N^2-1)} \tag{2}$$

In addition, there are evaluation indicators such as Kendall Rank Order Correlation Coefficient (KROCC) and Root Mean Square Error (RMSE).

In addition, Kendall Rank Order Correlation Coefficient (KROCC) and Root Mean Square Error (RMSE) is also the indicator of the IQA.

0.5 Traditional algorithm

0.5.1 FR-IQA

Peak Signal to Noise Ratio(PSNR) is the most widely used performance quantization method in the field of image and video processing. Usually used to evaluate the quality of an image after compression compared with the original image. The higher the PSNR, the smaller the distortion after compression: commonly involving support vector machines (SVM) and Bayesian methods.

$$PSNR = 10log_{10}(\frac{MAX^2}{MSE}), MSE = \frac{1}{MN}\sum_{i=1}^{N-1}\sum_{i=1}^{M-1}[I(i,j) - K(i,j)]^2 \tag{3}$$

0.5.2 RR-IQA

Although FR-IQA has achieved good results, in many applications, the image reference cannot be obtained, and only a part of the information or indirect characteristics of the reference image can be obtained, which has developed RR-IQA. The RR-IQA method provides a solution for situations where the reference image cannot be fully accessed. Maalouf et al [31] proposed an RR algorithm based on group transformation. Given a reference image and its distorted version, apply the image group to two images in order to extract the texture and gradient information of the image, and then filter this information through CSF And threshold processing to obtain the sensitivity coefficient, and finally compare the sensitivity coefficient of the distorted image with the sensitivity coefficient of the reference image to estimate the image quality. Guanawan et al. [32] proposed an RR-IQA algorithm that operates on blocked or fuzzy degraded images based on local harmonic analysis, calculates local harmonic amplitude information from edge detection images, and then uses this information with distorted images for Estimate the image quality.

0.5.3 NR-IQA

Human can judge the quality of distorted images without errors without reference images, but from a computer perspective, this task is quite challenging. The NR-IQA algorithm attempts to evaluate image quality without using a reference image. Most NR-IQA algorithms try to detect specific types of distortion, such as blur, block effects, various forms of noise, etc.

0.6 Deep learning

In recent years, deep learning has attracted the attention of researchers and has achieved great success in various computer vision tasks. Specifically, CNN has shown superior performance on many standard object recognition benchmarks. One of the advantages of CNN is that it can directly take the original image as input and integrate feature learning into the training process. CNN has a deep structure that can effectively learn complex mappings while requiring minimal domain knowledge. Here mainly introduces the use of deep learning to train NR-IQA.

Le Kang et al. [19] used 5-layer CNN to accurately predict NR-IQA. This method takes 32 * 32 image blocks, uses local normalization, combines global max pooling, min pooling, and

Relu non-linear activation layers, selects the SVR loss function, and uses momentum SGD to train the model

Weilong Hou et al. Also used deep learning algorithms for image quality evaluation. Using the BIQA framework, integrated image representation, NSS features, mixed features, classification, and posterior probability calculations are integrated into one function. The detailed features of the 3-level wavelet transform are used as input. The restricted Boltzmann machine RBM is used for training. Learning, then fine-tuning by back-propagation algorithm, and finally divided the prediction results into 5 levels

Ke Gu et al. [33] introduced a new deep learning based image quality index (DIQI) to evaluate the quality of unreferenced images. First convert the RGB image to the YIQ color space and extract 3000 features from it. Then use the L-BFGS algorithm to train a sparse autoencoder. The input data is a matrix of s × 3000, and s represents the number of training samples. Design a 3 Layer DNN, initialize the DNN using the autoencoder just trained, then calculate the output using a linear function, and finally use the back-propagation algorithm to fine-tune the weight of each layer of the DNN according to the loss function.

Implementation

0.6.1 Part 1. Re-construct SalGan

In the first part of our implementation, we need to re-construct the SalGAN neural network. As we introduced in the literature review part, the SalGan is a saliency map prediction neural network to help us extract ROI in the input image. And we reconstruct it by using by PyTorch library. We also use the pre-trained model to help us get the weights.

0.6.2 Part 2. Automatically segmentation algorithm

After we extract the saliency map by SalGan neural network, we are trying to segment the ROI automatically and use this ROI as the input for our IQA process.

First of all, we use OpenCV and NumPy library to help us achieve this goal. We transformation the saliency map into a gray image at the first by gray() function in the OpenCV library. Then we use another function to get the binary mask of this gray image. Then we use morphologyEx() function in the OpenCV library to do some morphological operations to the binary mask. After this process, we get the General shape of our ROI predicted by the saliency map. Then we filling the holes and details in the binary mask to make it more accurate when we trying to get its shape. In this step, we use erode() and dilate() function, Morphological corrosion and expansion were performed for 4 times respectively. After this process, we get the accurate binary mask shape of the ROI, then we find the minimum court to include this ROI shape by minAreaRect() function, and save this box in the NumPy array with its rotation angle. After that, we get the position of four vertexes of this rectangle, and we can use the max height and width to segment this ROI.

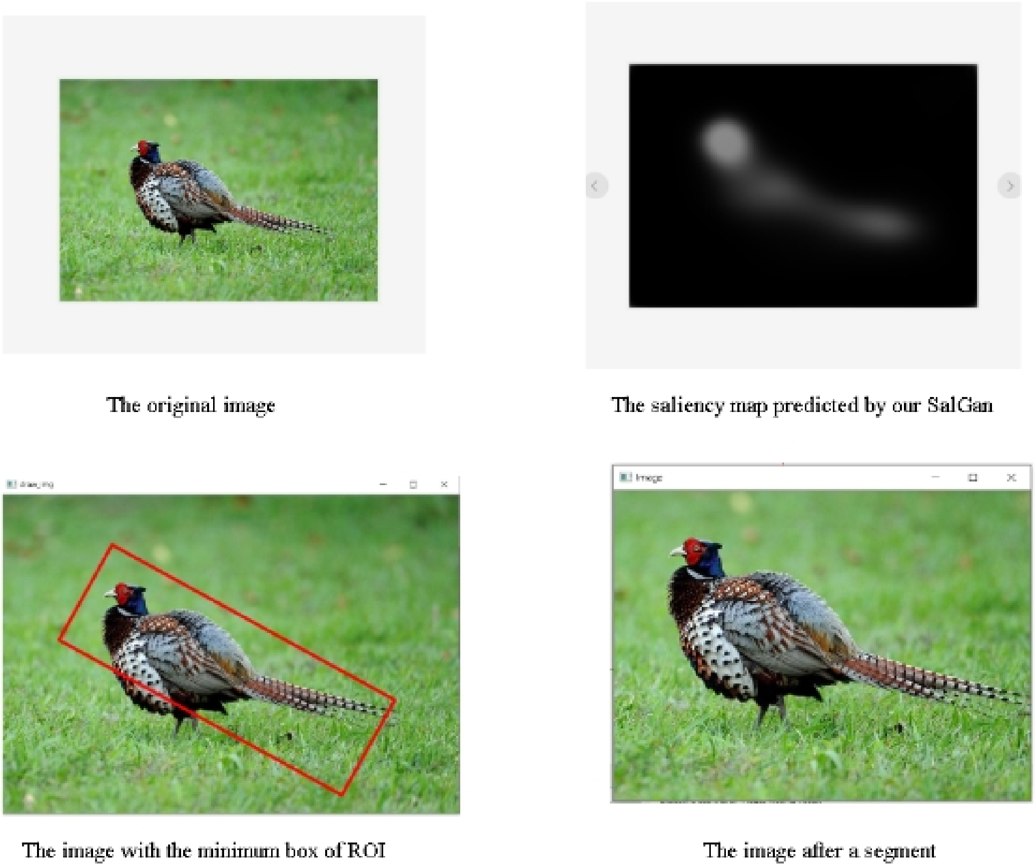

Figure 4: Example of SalGan and automatically segmentation algorithm

0.6.3 Part 3. IQA

In the IQA part, we are using the Natural image quality evaluator(NIQE) to score the original image and the image seperate by the segmentation algorithm, after we set the original image weight as 0.4 and the cut out image weight as 0.6 to get the total score of the quality of the image.

The correlation analysis results of the NIQE algorithm and the subjective evaluation show that the Pearson correlation coefficient is 0.910 1, the Spearman rank correlation coefficient is 0.946 1, and the average absolute value error is 0.016 0; compared to MSE, PSNR, SSIM, and MS-SSIM The four full-reference evaluation algorithms and the non-reference evaluation algorithm BIQI, NIQE algorithm show better prediction stability, monotonicity and consistency. Therefore, it can be seen that the NIQE algorithm has better consistency with the human subjective quality evaluation, closer to the human visual system, and can effectively carry out real-time image quality evaluation

In the example, we put the original image as input at the first, and we get scores of ROI and whole image separately. Then we put different weights into those two scores and get the final score of the original image which is 20.14. After that, we put some noise in the ROI of the original image then we get 16.98. At this step, we can see the noise points in the ROI reduces the quality of the original image significantly. We can get the same result easily by using human eyes. We can not see the building clearly in the image which makes it difficult to understand information contains in the image. Therefore the quality of the image is reduced.

Then we add some noise points into the background rather than ROI, we get 26.58 scores. As we opinion, the reason for this higher score is we put too much weight on the ROI part so it can easily change the final score of the image. And also the NIQE algorithm is not a really smart IQA algorithm so that it may give a higher score for a background with noise points. At least, we can conclude that noise points in the background do not really affect the quality of the image. We also can get the same conclusion by using human eyes, even there are some noise points in the background we can easily understand there is the sea in the background and we can see the building in the image clearly. It means we are not really affected by those noise points to understand the information in the image.

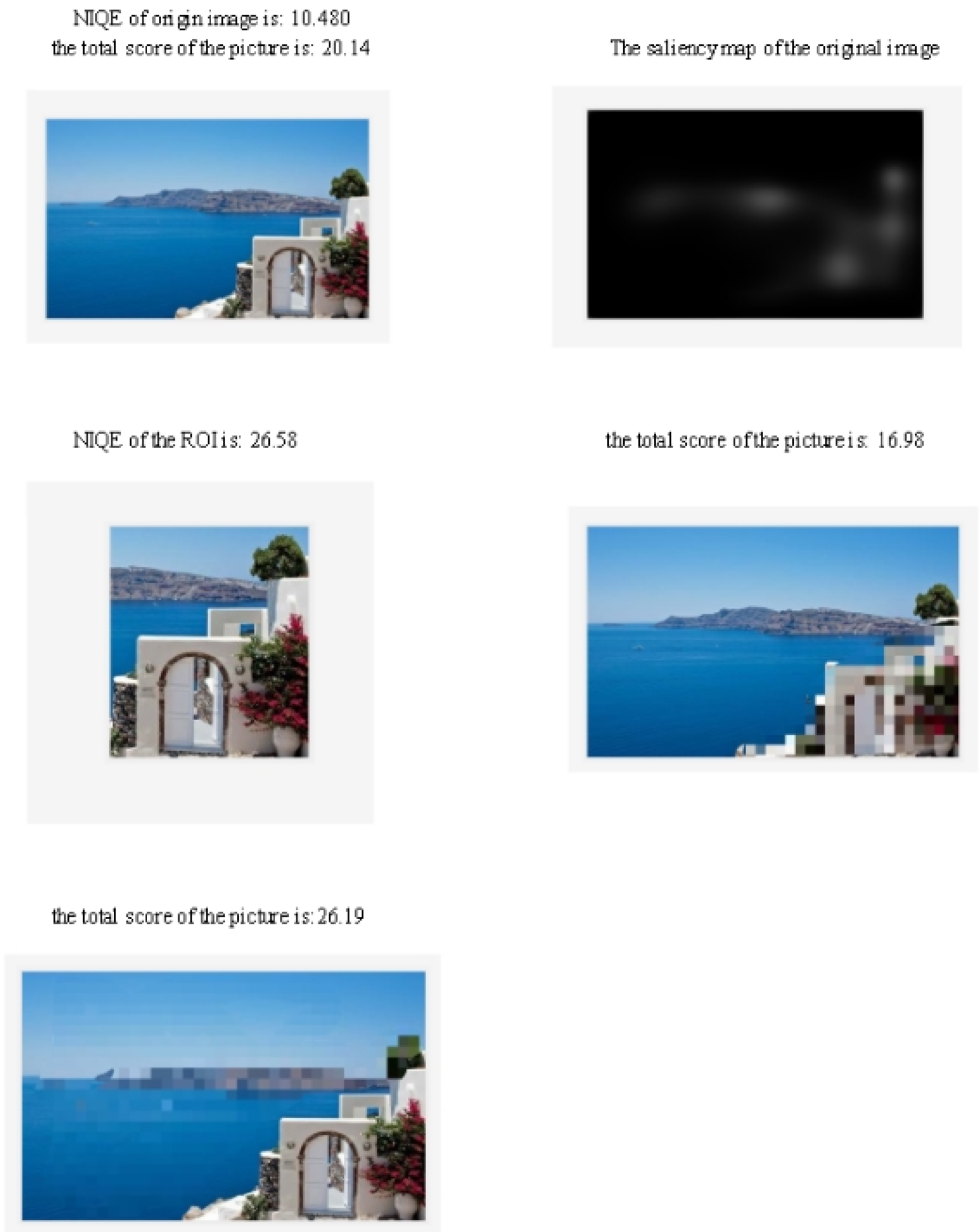

Figure 5: Example of the effect of ROI in IQA process

Conclusion

In this paper, we did not choose an IQA algorithm which uses a neural network, we choose an IQA algorithm which is not really advanced, but even in this type algorithm, the ROI also show a really important effect of IQA process. We believe that it will have a more important effect on other advanced algorithms and it can help to improve the accuracy of IQA score.

After experiments, we can say that the quality of ROI is playing an important role in the IQA process. Because ROI always shows the most important information about the image. And for the human, when we read an image, we want to understand the information contained in the image, this is the basic concept to do the quality assessment of an image. So we think that combine the ROI and IQA process provides a more accurate and understandable way to show the quality of the image. We hope in the future, we can combine the ROI with some IQA algorithm with the neural network to show a more accurate score.

References

[1] P. Le Callet and E. Niebur. Visual attention and applications in multimedia technologies. *Proceedings of the IEEE*, 101(9):2058–2067, Sep. 2013.

[2] Hani Alers, Judith A. Redi, Hantao Liu, and Ingrid Heynderickx. Studying the effect of optimizing image quality in salient regions at the expense of background content. *Journal of Electronic Imaging*, 22(4):1 – 11, 2013.

[3] H. Liu, U. Engelke, J. Wang, P. Le Callet, and I. Heynderickx. How does image content affect the added value of visual attention in objective image quality assessment? *IEEE Signal Processing Letters*, 20(4):355–358, April 2013.

[4] Xialei Liu, Joost van de Weijer, and Andrew D. Bagdanov. Rankiqa: Learning from rankings for no-reference image quality assessment. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[5] Alireza Alaei, Romain Raveaux, and Donatello Conte. Image quality assessment based on regions of interest. *Signal, Image and Video Processing*, 11:673–680, 2017.

[6] Dirk Walther, Laurent Itti, Maximilian Riesenhuber, Tomaso Poggio, and Christof Koch. Attentional selection for object recognition — a gentle way. In Heinrich H. Bülthoff, Christian Wallraven, Seong-Whan Lee, and Tomaso A. Poggio, editors, *Biologically Motivated Computer Vision*, pages 472–479, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[7] B. Sch?lkopf, J. Platt, and T. Hofmann. *Graph-Based Visual Saliency*, pages 545–552. MITP, 2007.

[8] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2106–2113, Sep. 2009.

[9] Eleonora Vig, Michael Dorr, and David D. Cox. Large-scale optimization of hierarchical features for saliency prediction in natural images. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2798–2805, 2014.

[10] Junting Pan, Cristian Canton Ferrer, Kevin McGuinness, Noel E. O'Connor, Jordi Torres, Elisa Sayrol, and Xavier Giro i Nieto. Salgan: Visual saliency prediction with generative adversarial networks, 2017.

[11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 2016.

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[13] Zhou 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*, 13(4):600–612, April 2004.

[14] A. Rehman and Z. Wang. Reduced-reference image quality assessment by structural similarity estimation. *IEEE Transactions on Image Processing*, 21(8):3378–3389, Aug 2012.

[15] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, Dec 2012.

[16] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai. No-reference quality assessment of contrast-distorted images based on natural scene statistics. *IEEE Signal Processing Letters*, 22(7):838–842, July 2015.

[17] Q. Li, W. Lin, J. Xu, and Y. Fang. Blind image quality assessment using statistical structural and luminance features. *IEEE Transactions on Multimedia*, 18(12):2457–2469, Dec 2016.

[18] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin. No reference quality assessment for screen content images with both local and global feature representation. *IEEE Transactions on Image Processing*, 27(4):1600–1610, April 2018.

[19] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[20] W. Hou, X. Gao, D. Tao, and X. Li. Blind image quality assessment via deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 26(6):1275–1286, June 2015.

[21] K Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. pages 1–14. Computational and Biological Learning Society, 2015.

[22] S. Bosse, D. Maniry, K. Muller, T. Wiegand, and W. Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, Jan 2018.

[23] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 27(3):1202–1213, March 2018.

[24] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, May 2010.

[25] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, March 2013.

[26] L. Zhang, L. Zhang, and A. C. Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, Aug 2015.

[27] Wufeng Xue, Lei Zhang, and Xuanqin Mou. Learning without human scores for blind image quality assessment. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[28] Y. Zhang and D. M. Chandler. Opinion-unaware blind quality assessment of multiply and singly distorted images via distortion parameter estimation. *IEEE Transactions on Image Processing*, 27(11):5433–5448, Nov 2018.

[29] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao. dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Transactions on Image Processing*, 26(8):3951–3964, Aug 2017.

[30] P. Ye, J. Kumar, L. Kang, and D. Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1098–1105, June 2012.

[31] A. Maalouf, M. Larabi, and C. Fernandez-Maloigne. A grouplet-based reduced reference image quality assessment. In *2009 International Workshop on Quality of Multimedia Experience*, pages 59–63, July 2009.

[32] Irwan Prasetya Gunawan and Mohammed Ghanbari. Reduced-reference picture quality estimation by using local harmonic amplitude information.

[33] K. Gu, G. Zhai, X. Yang, and W. Zhang. Deep learning network for blind image quality assessment. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 511–515, Oct 2014.