2019 International Conference on Aviation Safety and Information Technology (ICASIT 2019)

A Survey of Non-reference Image Quality Assessment Based on CNN

Yao Ma
Liaoning University of Technology
Jinzhou, China

Fuming Sun
Liaoning University of Technology
Jinzhou, China

Shijie Hao
Hefei University of Technology
Hefei, China

***Abstract*—Due to the serious shortage of training data, CNN's research on non-reference image quality assessment (NR-IQA) is very constrained. In this paper, the existing neural network research on NR-IQA is summarized, and the methods to solve the problem are divided into three types: image segmentation, pre-training migration and unsupervised sequence learning. Then from these three aspects, a more representative algorithm is selected to test and compare performance on different datasets, and analyze the advantages and problems of the three. In the end, the direction of NR-IQA development is discussed, which provides a comprehensive reference for researchers in this field.**

Keywords—CNN, image segmentation, pre-training migration, unsupervised sequence learning, NR-IQA

I. INTRODUCTION

Image quality assessment is a crucial part of the images transmission process. Existing methods can fall into three categories: full reference (FR-IQA), reduced reference (RR-IQA), and no reference (NR-IQA). FR-IQA means that all the information of the pristine undistorted image is known, studying from the difference between the distorted image and the pristine image, such as several classic algorithms: SSIM [1], GMSD [2], FSIM [3] and so on. RR-IQA means that only some of the features of the pristine image are known as a reference, and NR-IQA is evaluated without reference information at all. Because the pristine image is often difficult to obtain in practical applications, the researchers pay more attention to the non-reference image quality assessment. The existing NR-IQA algorithms can be roughly divided into two categories: methods based on natural statistical characteristics and methods based on learning training.

The former study unfolds the statistical properties of distorted images and pristine undistorted images [4,5,6]. For example, BRISQUE [4] shows that the statistical property distribution of natural images approximates the generalized Gaussian distribution (GGD), which is warped by distortion. Therefore, the statistical characteristics on the GGD spectrum are used to distinguish the types of distortions and to evaluate the image distortion quality. According to the statistical data of structural features, BLIINDS-II [5] believes that the degree of distortion and the type of distortion change the DCT coefficients, so features are extracted in the DCT domain to predict the quality score. The latter study uses feature learning instead of hand-craft feature extraction. For example, CORNIA [7] proves the validity of using the pristine image directly as input, and uses k-means to encode local features and support vector machine to quality evaluation.

The former research relies heavily on the distortion type and local feature calculation, making it very limited. The latter, due to CNN, has made the algorithm get rid of these limitations, and the performance has been greatly improved. However, the effectiveness of the CNN network is very dependent on the data. Existing image quality assessment data sets are not sufficient to support training network models with a large number of parameters. The LIVE datasets contains a total of 779 distortion images for 5 distortion types; The TID2008 datasets has a total of 1700 distortion images for 17 types of distortion; The TID2013 datasets has a total of 3,000 distortion images for 24 distortion types. As the type of distortion increases, the number of distorted images for each distortion type in the datasets does not increase. In order to solve this problem, CNN's research on NR-IQA can be roughly divided into three categories: image segmentation, pre-training migration and unsupervised sequence learning.

II. IMAGE SEGMENTATION PROCESSING

Image segmentation processing refers to the method of dividing the entire image into several image patches to expand the datasets. There are a variety of ways to crop, such as no overlapping cropping, overlapping cropping, or random cropping. Extending the datasets by segmenting the image is the most direct and effective way, but it also brings two problems. Firstly, in the datasets, only the quality of the entire image is included, and there is no quality score for the image patch after the segmentation. Therefore, how to obtain the validly quality score of image patch becomes a very important research branch. Secondly, if you take a small image patch as input, the resulting output is naturally the quality score of the image patch. It is also important to combine the scores of the cropped image patches into the quality scores of the entire image. By different acquisition methods, we will discuss two aspects.

A. Directly Given

The easiest way is to give the score of the image directly to the image patch, which is cropped from the image. The more classic is the work of Kang et al. [8], simple and straightforward and the results are good. He splits the entire image into 32*32 image patches without overlapping, and the scores of these image patches are directly equivalent to the scores of the image. The supervised training with the expanded datasets and the results obtained in a very simple six-layer network were very competitive at the time, which fully proved the advantages of CNN in NR-IQA.

However, there will be noise when the label is directly applied, and the quality score of the image patches will be distorted, which will have a great impact on the accuracy of the network training. Therefore, Lu Peng et al. [9] proposed an entropy-based method to analyze and verify the influence of image entropy on the distortion of the image, and use information entropy as the weight of the loss function to improve the accuracy.

Even if you get a quality score directly given, considering the type of distortion in your network may increase its accuracy. After image patches quality score processing, Fan et al. [10] adopts a method of combining multiple types of distortions, inputting image patches into multiple networks, and obtaining multiple quality scores relative to each type of distortion.

B. Combined FR-IQA

The so-called combined FR-IQA usually refers to the method of obtaining a quality score of an image patch using a full reference method. After years of research by researchers, FR-IQA has been sufficiently accurate, so combining it is a very effective way. There are many existing effective FR-IQA algorithms. How to effectively select and combine them has a great impact on the performance of the network. Wen et al. [11] considered the methods of Mean Square Error (MSE) and SSIM respectively, and proposed a specific formula for calculating the quality score of image patches, which was used as a label for supervision training.

Kim et al. [12] combined four classic FR algorithms: SSIM, GMSD, FSIM, VSI. Through these FR-IQA algorithms, he calculates a local quality map of the image with a quality score and uses it to train the network. The combination of FR-IQA algorithms is not as good as possible, and the combination with the network is the key. Bare et al. [13] directly adopted FSIM as a label, combined with partial residual knowledge, and added two layers of sum to construct a deep CNN model. Performance has indeed improved without pre-training.

C. Pooling

There are two problems after image segmentation, the label problem from the entire image to the image patch and the pooling problem from the image patch to the entire image. There are many solutions to the former, which we have explained in detail, and the latter also has a variety of treatments, such as average pooling, max-pooling, min-pooling, feature connections and so on. The different choices of pooling will also have a certain impact on the accuracy of the entire network.

Perhaps just changing the pooling method, the performance also will be improved. The best way to discriminate is to make the pooling method as the only variable. For example, Bianco et al. [14] compared the different pooling methods of the same network. When the conditions of cropping images, network structure, and pre-training are the same, the results are compared for different pooling methods: feature fusion (average pooling of image patch features), feature connectivity, and prediction result fusion (average pooling of quality scores). The results show that the first pooling method works best. It can be seen that choosing the right pooling method has a certain effect on the performance improvement.

III. Pre-Training Migration

Although image segmentation does greatly improve the problem of insufficient datasets, a large amount of preprocessing makes the end-to-end superiority of CNN lost. So, with the development of deep learning, migration learning has gradually become popular. Through similar task migration, the network already has certain recognition ability, and then fine-tuning is performed by using a small datasets of NR-IQA. This eliminates the image patch quality score noise situation and the pooling error problem, at the same time, improves the ability of generalization.

In NR-IQA, most of the existing migration learning considers it a classification task. The type of distortion or the degree of distortion is used as a classification target to fine-tuning the model. For example, according to the type of distortion, Gao et al. [15] used datasets with size C*840*5 for pre-training (where C is the number of distortion types) to first determine the distortion type of the image. He then considers the effect of the distortion type on the quality score, and builds a loss function that contains the predicted distortion type results. There are many ways to migrate, or you can use only part of the structure of the pre-trained network.

Choosing the right network is critical during the migration process. For example, Kim et al. [16] compared the results of different networks through the same pre-training. The AlexNet and ResNet networks were tested separately. The experimental results show that the results of the ResNet network are better. Overall, the migrated network performs better on the Live Challenge dataset than other networks. The author believes that this datasets is a real distortion picture, while other datasets are artificially synthesized distortion images, and the models pre-trained by real images have stronger feature extraction capabilities for real images.

The results of different pre-training methods for the same network may also be different. Bianco et al. [14] compared the results of the three pre-training methods. ImageNet, Places and ImageNet+Places are pre-trained on Caffe-Net. The results show that the more the CNN recognition ability training, the more effective the network is for feature extraction, which indicates the overall content of the image.

It can be seen that after pre-trained migration, the generalization ability is improved, the perception of real pictures is enhanced, and the tasks across data sets can be better accomplished. There is a big improvement in the Live Challenge dataset that has struggled with many network models. But to be more competitive, the amount of data is still a critical issue. It may be a good idea to combine the methods of image segmentation to further expand the data set so that the network is fully trained.

IV. Unsupervised Sequence Learning

Because artificial subjective scores are no longer used, datasets can be extended from only specific datasets to a wide variety of readily available datasets, or to artificially generate distorted image datasets without quality scores. This method greatly increases the data of the training image, thereby solving the problem of insufficient data volume. The dataset used by Ma et al. [17] is a 840 pristine undistorted image collected in the real world, which is subjected to four types of five levels of distortion, and constitutes 80 million discriminable image pairs.

Even for unsupervised learning, it can be fine-tuned after network training. For example, Liu et al. [18] used the idea of Siamese network [19] to train Vgg-16 with a number of known hierarchical order image pairs. Compared to traditional networks, the network of image pairs has two network branches that share weight, and the output of the last layer is a scalar. After using the image pair to train the Siamese network, Liu selects one of the branches and uses the IQA datasets to fine-tune it to form the CNN that ultimately implements NR-IQA.

Moreover, sequence learning is not only used for image pair ranking, but also for image patch quality score acquisition. For example, Ye et al. [20] used a non-supervised hierarchical fusion method to perform a consistent ranking of several FR methods and obtained a quality score RRFscore(Ii), which in turn is adjusted by the selected FR method to make it a valid score.

Since it is not limited by the existing IQA datasets, unsupervised sequence learning does not show much fluctuations when performing performance tests on different datasets. However, the models trained by this method are often not very targeted and have a poor consistency with the subjective quality scores, so the overall performance may not be as competitive as the former two. Taking full advantage of its stability and strengthening some consistency training may make the unsupervised sequence learning method more effective. The work of Ye et al. is a good start.

V. EXPERIMENT

SROCC and PLCC are two very important indicators to consider the performance of image quality assessment.

The SROCC considers the rank correlation of the two sets of data, assuming that there are two sets of data for X and Y, which are the predicted scores and the ground labels in the image quality assessment. SROCC requires that the two sets of data be sorted from big to small, the lower the score, the higher the level. Calculate the difference between the predicted score and the real label of each input image, and find the square, shown in Equation (1).

$$\mathrm{SROCC} = 1 - 6\sum_{i=1}^{n} d_i^2 / n(n^2 - 1) \qquad (1)$$

SROCC focuses on the relative amount of data, so a monotonic nonlinear change to the set of data does not affect its results. PLCC considers the linear correlation coefficient between the two sets of data. The knowledge of the covariance and standard deviation product of probability theory is used. The formula is shown in Equation (2).

$$\mathrm{PLCC} = \mathrm{COV}(X, Y) / \delta_X \, \delta_Y \qquad (2)$$

LIVE is one of the earliest image evaluation datasets. It contains the first five types of distortion, and there are more than one hundred distortion images for each type of distortion. A lot of research is based on this datasets. We have listed some of the results of training and testing on the LIVE dataset, as shown in Table I.

In Table I, several typical algorithms are selected for comparison in three different methods. The first two columns is the result of train and test on LIVE datasets. The first row and the second row are algorithms directly given after image segmentation, the third row is a combination of FR. It can be concluded that for the quality score acquisition of the image patch, relying on the reliable and effective FR method can indeed achieve better results. And the fourth row and fifth row are pre-training migration algorithms. As can be seen from Table I, the results of the migrated network is closely related to the selected network. The last row are unsupervised sequence learning. In comparison, the results are similar to the former and do not have a strong advantage. The last two columns is the result of train and test on TID2008 datasets. It can be seen that the experimental results of either SROCC or PLCC have a significant decline in the datasets.

TABLEI. TRAIN - TEST ON THE LIVE OR TID2008 DATASET

	SROCC	PLCC	SROCC	PLCC
CNN[1]	0.956	0.953	0.920	0.903
DVRM-S [7]	0.937	0.942	0.916	0.904
BIECON[4]	0.961	0.962	0.923	-
ALE-F[15]	0.947	0.952	-	-
DEEPBIQ[2]	0.98	0.97	0.950	0.950
DIPIQ[3]	0.958	0.957	0.877	0.894

Note: DIPIQ is the training and testing done in the TID2013 datasets.

TABLEII. TRAIN ON THE LIVE DATASET - TEST ON THE TID2008 DATASET

SROCC	JP2K	JPEG	WN	BLUR	ALL4
DVRM-S[7]	0.943	0.930	0.909	0.773	0.894
DVRM-A[7]	0.922	0.926	0.909	0.764	0.895
BLISS-C[12]	0.923	0.926	0.807	0.880	0.899
BIECON[4]	0.878	0.941	0.842	0.913	0.923
DIPIQ[3]	0.926	0.932	0.905	0.922	0.877
PLCC	JP2K	JPEG	WN	BLUR	ALL4
DVRM-S[7]	0.945	0.942	0.919	0.801	0.911
DVRM-A[7]	0.893	0.940	0.925	0.828	0.896
BLISS-C[12]	0.941	0.952	0.770	0.880	0.917
DIPIQ[3]	0.948	0.973	0.906	0.928	0.894

Table II is the test result of several algorithms across data. Compared with the results trained in the TID2008 dataset, whether it is an algorithm such as image segmentation processing or a pre-training migration algorithm, the results are further reduced on the basis of the original. It can be seen that in the test of cross-datasets, the network performance degradation using image segmentation processing is the most obvious; Although the pre-trained CNN has a certain generalization ability, the result is only a relatively small decrease in performance, but the effect cannot be avoided; Unsupervised sequence learning, performance has not changed much, because it does not rely on training with datasets, but when training tests on the same dataset, the results are worse than the other two.

It can be seen that the generalization ability of the current non-reference image quality assessment algorithm needs to be improved. Many networks that perform well on LIVE datasets are not very satisfactory on other datasets. Even if they are retrained, it is difficult to achieve the same level.

What's more, instead of using other datasets training, it is directly tested on it, and the results are significantly reduced. With the development of the times, the type of image distortion is constantly increasing, and the generalization ability of the network not only requires the ability to cross datasets, but also requires able to exploit the type of distortion that has been trained to predict the type of distortion that has not been trained. Most of the current research focuses on the five types of distortions included in the LIVE dataset, without fully utilizing the content extended by subsequent datasets. Instead of focusing on several types of distortion studies, research on the degree of distortion, or even a mixture of distortion images of multiple distortion types, will be the focus of future reference image quality assessment.

VI. Conclusion

Quality evaluation of images is a crucial part of the image transmission process. Non-reference image quality assessment is a typical machine learning problem in terms of feature extraction and recognition, and it can be found that CNN-based methods do improve performance through existing research. How to use CNN to solve existing problems will be a research hotspot in this field.

Acknowledgments

This work was partially funded by the Natural Science Foundation of China under Grant 61772171.

References

[1] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. ImageProcess., vol. 13, no. 4, pp. 600–612, Apr. 2004

[2] W. Xue, L. Zhang, X. Mou, and A. Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” IEEE Trans. Image Process., vol. 23, no. 2, pp. 684–695, Feb. 2014.

[3] L. Zhang, D. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” IEEE Trans. Image Process. vol. 20,no. 8, pp. 2378–2386, Aug. 2011.

[4] Mittal A, Moorthy A K, Bovik A C. No-Reference Image Quality Assessment in the Spatial Domain. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2012, 21(12):4695.

[5] M. Saad, A. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the DCT domain.IEEE Transactions on Image Processing, 21(8):3339–3352, Aug. 2012.

[6] Gu J, Meng G, Wang L, et al. Learning deep vector regression model for no-reference image quality assessment. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2017:2961-2965.

[7] Gu J, Meng G, Wang L, et al. Learning deep vector regression model for no-reference image quality assessment. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2017:2961-2965.

[8] Kang L, Ye P, Li Y, et al. Convolutional Neural Networks for No-Reference Image Quality Assessment.IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:1733-1740.

[9] Peng Lu, Genqiao Lin, Guoliang Zou. et al. Study on non-reference image quality evaluation methods based on information entropy and deep learning. Computer application research, 2018(12):1-7.

[10] Fan C, Zhang Y, Feng L, et al. No Reference Image Quality Assessment based on Multi-expert Convolutional Neural Networks. IEEE Access, 2018, PP(99):1-1.

[11] Wen H, Jiang T. From image quality to patch quality: An Image-Patch Model for No-Reference image quality assessment// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2017:1238-1242.

[12] Kim J, Lee S. Fully Deep Blind Image Quality Predictor. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(1):206-220.

[13] Hou W, Gao X, Tao D, et al. Blind image quality assessment via deep learning. IEEE Transactions on Neural Networks & Learning Systems, 2017, 26(6):1275-1286.

[14] Bianco S, Celona L, Napoletano P, et al. On the use of deep learning for blind image quality assessment. Signal Image & Video Processing, 2016(3):1-8.

[15] Kim J, Zeng H, Ghadiyaram D, et al. Deep Convolutional Neural Models for Picture-Quality Prediction: Challenges and Solutions to Data-Driven Image Quality Assessment. IEEE Signal Processing Magazine, 2017, 34(6):130-141.

[16] Gao F, Wang Y, Li P, et al. DeepSim: Deep similarity for image quality assessment. Neurocomputing, 2017.

[17] Ma K, Liu W, Liu T, et al. dipIQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs. IEEE Trans Image Process, 2017, 26(8):3951-3964.

[18] Liu X, Weijer J V D, Bagdanov A D. RankIQA: Learning from Rankings for No-Reference Image Quality Assessment.// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017:1040-1049.

[19] Chopra S, Hadsell R, Lecun Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005:539-546 vol. 1.

[20] Ye P, Kumar J, Doermann D. Beyond Human Opinion Scores: Blind Image Quality Assessment Based on Synthetic Scores.// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:4241-4248