Tech Science Press

No-Reference Stereo Image Quality Assessment Based on Transfer Learning

Lixiu Wu[1,*], Song Wang[2] and Qingbing Sang[3]

[1]Jiangsu Key Construction Laboratory of IoT Application Technology (Wuxi Taihu University), Wuxi, 214000, China
[2]Pactera Yuanhui Technology (Wuxi) Co., LTD., Wuxi, 214000, China
[3]Jiangnan University, Wuxi, 214000, China
*Corresponding Author: Lixiu Wu. Email: wlx9009@foxmail.com
Received: 12 January 2022; Accepted: 26 April 2022

Abstract: In order to apply the deep learning to the stereo image quality evaluation, two problems need to be solved: The first one is that we have a bit of training samples, another is how to input the dimensional image's left view or right view. In this paper, we transfer the 2D image quality evaluation model to the stereo image quality evaluation, and this method solves the first problem; use the method of principal component analysis is used to fuse the left and right views into an input image in order to solve the second problem. At the same time, the input image is preprocessed by phase congruency transformation, which further improves the performance of the algorithm. The structure of the deep convolution neural network consists of four convolution layers and three maximum pooling layers and two fully connected layers. The experimental results on LIVE3D image database show that the prediction quality score of the model is in good agreement with the subjective evaluation value.

Keywords: No-reference; stereo image quality assessment; convolution neural network; transfer learning; phase congruency transformation; image fusion

1 Introduction

In 2009, the release of 3D movie Avatar brought a wave of movie watching across the world, creating a new record in the history of movie taking USD28 billion at the box-office globally, which amazed the world, according to the statistics of Motion Picture Association of America, 3D movie mushroomed by 50% each year, currently, more than half of movies provide 3D copies. Starting 2010, 3D of several brands had been launched in the mainland China in succession. On 1st January 2012, CCTV cooperated with many television sets in our country, jointly rolling out the first 3D trial channel of China. The reason why the 3D picture and video are so popular is mainly because the production of 3D videos applies the principle that the parallax angle that people use two eyes to observe objects can tell the distance, then stereoscopic vision is generated, which makes the audiences have the feelings that they are in the scene personally. Recently, 3D technology has been the foundation of innovation and new vantage point of product design, manufacture, management, marketing, consumption, and

so on, 3D three-dimensional picture and video technique have been the trendy research field, among them include quality assessment of 3D image.

The algorithm of three-dimensional image quality assessment can be divided into three types according to degree of dependence that distorted image to original image: Full Reference (FR), Reduced Reference (RR) and No Reference (NR) image quality assessment. Some image quality database and subjective mean opinion score (MOS) are republished in accordance with a serial of suggestions of video quality expert group [1], meanwhile, the image evaluation methods based on various effective models are brought out. In the past, peak signal to noise ratio (PRNR) was deemed as the best image quality assessment method before structural similarity (SSIM) [2] was emerged, on the assumption that human visual perception system is easy to extract structure information of image, the document [3] verify SSIM can have better performance than traditional PSNR on the LIVE database. At present, the research on the flat image quality evaluation has made significant achievement, but the research regarding stereoscopic image quality evaluation has certain challenge. In the method of FR image quality assessment, such as the document [4–7], in which, You et al. [4] and others use the method of FR image quality assessment to the method of flat image quality assessment directly in the evaluation of stereoscopic image, but this kind of method cannot take the differences between 3D image and flat image into consideration, which cannot fully imitate the complicate vision perception mechanism of human, performance of evaluation is below the average. The document [6] puts forward one type of FR stereo picture quality assessment model on the foundation of binocular energy quality metric (BEQM) by studying the binocular competence, the influence of binocular visual perception on the stereo image quality assessment is considered when they are producing this model, so it is consistent with the subjective evaluation of stereo image, but it is not sufficient for this model to know the binocular visual perception, and this model is relatively complicated, which is not suitable for NR stereo image quality assessment. The document [7] puts forward a method of FR quality assessment method based on colorful stereo image, due to the complex of human vision system, this method has not had a reasonable vision model, in addition, FR stereo image quality assessment need gain all the information of reference image, so this method has certain limitation. The document [8] No reference quality assessment for stereoscopic images by statistical features. No-reference Stereoscopic Image Quality Assessment Based on Binocular Fusion and Binocular Rivalry [9] and No-reference stereoscopic image quality assessment based on deep feature learning [10]. The document [11] put forwards a no-reference stereoscopic image quality assessment method based on wavelet transform. The document [12] put forwards a method of NR stereo image quality assessment based on in-depth learning, which firstly left and right views by Gabor filter, the statistical features of different sizes and directions as the monocular, then according to the binocular competence feature of HVS, merge the left and right views, extract the histogram with the fusion of direction and gradient as the binocular feature, through deep belief network (DBN) to train regression model, and at last, predict the quality scores of left view and right views and gain the quality score of the whole image by combining the above two scores.

Deep learning has been succeeded in image identification, voice processing, understanding of natural language, and other fields since the Hilton and Salakhutdinow were proposed in 2006. Using the deep learning can solve the problem of stereo image quality assessment, but the biggest obstacle lies on the mass data required in the course of training models. The requirement of such a large number of data is because the machine will encounter enormous parameter in the learning. However, the relation gained in one model training targeting at certain type of data can easily apply to different problems in the same field, which can be called transfer learning [13]. This article is to transfer the quality

assessment models of the flat image to the ones of stereo image, and achieved a great assessment indicator.

2 Image Fusion and Phase Congruency

2.1 Image Fusion

Image fusion is to merge complementary information of two and more than two images into one new image, fitting the visual perception of human more and more convenient for the image to have further processing and analysis. This article applies image fusion algorithm [14] of PCA (principal component analysis), which is a multidimensional orthogonal linear transformation based on statistical feature. In the fusion, firstly take the left and right views of stereo pictures as the matrix input, then transfer into one-dimensional vector X, Y, count covariance matrix of two according to Eq. (1), secondly, take count the eigenvalue and eigenvector of covariance matrix, finally by comparing the eigenvalue, confirm the fusion coefficient of image according to eigenvector, according to Eq. (2),

$$Cov\,(X,\,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \tag{1}$$

$$Cylopean = a * left\_image + b * right\_image \tag{2}$$

In the formula, X, Y respectively represent serialization vector of left and right views, n represents the length of vector.

According to the above calculation, select a random image pair from LIVE3D image library, carry out the left and right views with PCA fusion processing, the effect is shown in the Fig. 1:

(a) left views

PCA fusion

(c) fusion image
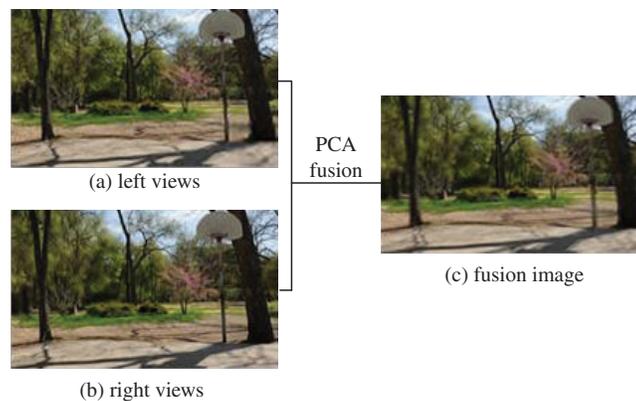
(b) right views

Figure 1: Schematic diagram of PCA fusion about left views and right views in stereoscopic image

We can see from (a), (b), (c) in the Fig 1, the image after PCA fusion is a little vague and double image visually, and other problems, this is because in the fusion of image, it not only collects the flat information of stereo image, but also collects the in-depth information specific for stereo pictures. For the convenience of extracting features of images, this article the left and right views of stereo image are combined into one image by PCA fusion method.

2.2 Phase Congruency

Phase Congruency (PC) is raised by Morrone et al. [15] and more people when they find that the image features are always happened at the big overlapping area in the research of extracting the image feature. This model is not affected by the change of contrast ratio of the image, and can have better image visual perception when the luminance change, which is more suitable for the visual mechanism of human. Its application in the image quality assessment also works well, such as in the document [16,17], they use the PC in the image quality assessment, among them, [16] is first to conduct PC conversion to distorted images, and get the images of PC, and images of PC with the largest and smallest covariance, then calculate the gradient entropy of these three images and refer to the gradient entropy and gradient mean value of images, finally through the training of GRNN, predict the quality score of images; the document [17] is first to measure the distorted degree of distorted images in the similarity of local subdomain through the PC of distorted images and reference images, use the mean value and extreme value of PC of images in each subdomain as the weighted coefficient PC of images, calculate the weighted mean of relevant coefficients of each subdomain, and get the prediction quality of images.

Given $I(x)$ is the one-dimensional signal, M_n^e and M_n^o respectively represent even symmetric filter and odd symmetric filter of n scale, the convolution of it and I can represent:

$$[e_n(x), o_n(x)] = \left[I(x) * M_n^e, I(x) * M_n^o\right] \tag{3}$$

$$F(x) = \sum e_n(x), H(x) = \sum o_n(x) \tag{4}$$

$$\sum_n A_n(x) = \sum_n \sqrt{e_n(x)^2 + o_n(x)^2} \tag{5}$$

So the calculated mode of PC of one-dimensional signal is as follows:

$$PC(X) = \frac{\sqrt{F^2(x) + H^2(x)}}{\varepsilon + \sum_n A_n(x)} \tag{6}$$

In it: ε is a relatively small positive number, to avoid the denominator to be zero, $H(x)$ is one-dimensional Hilbert transformation of $F(x)$, it can also be used in the two-dimensional ones, the calculated mode of PC is as follows:

$$PC(x) = \frac{\sum_o |E_o(x) - T_o|}{\sum_o \sum_n A_{no}(x) + \varepsilon} \tag{7}$$

In the $E(x) = \sqrt{F^2(x) + H^2(x)}$; T_o represents noise compensating factor; ε represents a very small positive number to prevent the denominator from being zero.

This article selects randomly one reference image and one distorted image from LIVE3D image library, we can see PC can shake off the not important background information in the image the effect is shown in the Fig. 2, remain the structure features of images well, it is easy to discover that structure outline of (d) is more complicated than that of (b) from comparing (b) PC of reference image with (d) PC of distorted image, with the reason that there are various distorted elements in the distorted images, such as blur, noise, JPEG and JPEG2K compression, etc.

(a) Reference image   (b) PC of reference image

(c) Distorted image   (d) PC of distorted image
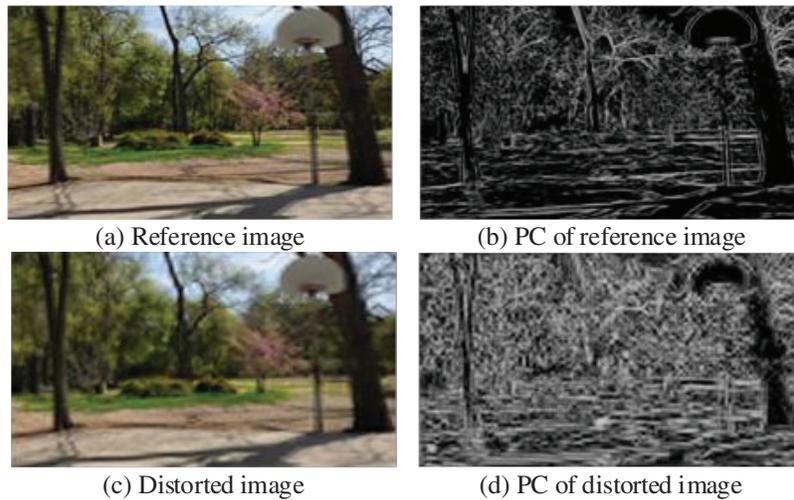
Figure 2: Phase congruency of reference image and distortion image

3 Based on the Internet Frame of Transfer Learning

3.1 Transferring Learning

Transfer learning is to learn new field by using the knowledge you already had. The knowledge used for transferring mainly include the transfer of data, feature and model parametric. Among them, data transfer is to extract the data suitable for task two in the task one, and combine the other data aiming at task two and one to get training; the feature transfer is to use the same features in the old task and new task to the training and learning of new task; model parameter transfer, similar to feature transfer, which not only transfers the feature into the similar tasks, but also the model can be transferred, such as the document [11] transfer ImageNet network structure. Transfer learning can put the trained model in the small new dataset and similar tasks, which can be used in small-scale dataset, and reduce the training time as well, mostly important, it works well.

LIVE3D stereo picture dataset has 725 pairs of distorted images, which are typical small training samples, it is not easy to have good performance index with these samples to train deep Convolutional Neural Network (CNN). The lab used the expanded 80 thousand flat distorted image a few days ago to train a flat quality evaluation model of good performance in deep CNN. This article transfers this flat model in the stereo picture quality evaluation using the technique of transfer learning, continuing to train stereoscopic distorted image on the trained model parameter, change of the Solver parameters test_iter from 1000 to 500, learning rate base_lr change from 0.01 to 0.001.

3.2 Network Architecture of CNN (Convolutional Neural Network)

CNN is one of widely used models in deep learning, which is developed from multilayer percep-tron, designing for image identification and other problems originally. But now CNN is used widely on the image and video fields, and sound signal, textual data, etc. CNN normally is comprised of several convolutions, and the image extract the local features of each image in convolution through filters of different convolution cores and bias, every convolution core can reflect a new image, then output result of filter in the convolution cores are activated by non-linear activation function, finally the pooling function are done for the result of activation function (i.e., down sampling). In most practical circumstances, choosing the features by hand is not stable to a large extent, and waste time

and energy. CNN just overcome these weaknesses, the weight of CNN is similar to the traditional BP neural network, all of them use backward propagation, so CNN can be directly to deal with input image, avoiding complicated preprocessing of the images. The big advantage of CNN is the sharing of weights, shrinking the parameter of neural network, which can prevent overfitting and makes the model of neural network easier.

This article uses the Caffe frame [18], refers to AlexNet [19] network architecture, uses the model parameter that already trained on the flat library, the network models transferred in the article is a convolution neural network containing 4 convolution layers and 2 fully connected layers (FC layer), and use built-in Euclidean Loss Layer in Caffe to solve the regression problem the effect is shown in the Fig. 3. Euclidean loss calculation methods are seen in Eq. (8):

$$E_{loss} = \frac{1}{2N} \sum_{n=1}^{N} \| Y_n - y_n \|_2^2 \tag{8}$$
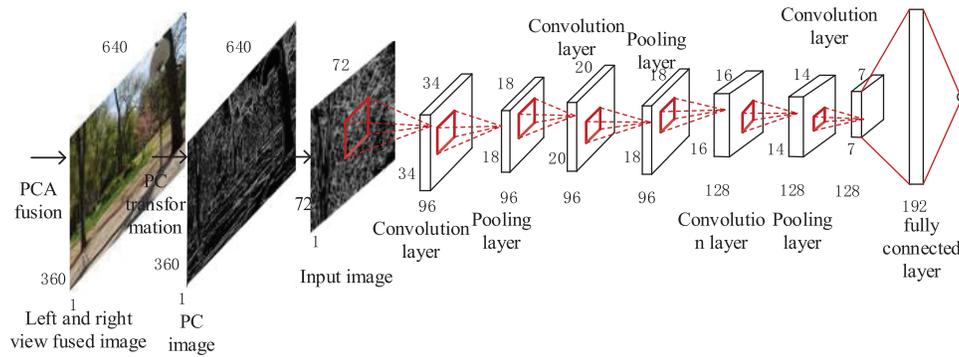
Figure 3: The network architecture

In it, E_{loss} represents loss value, Y_n represents prediction values of $(N * C * H * W)$ as the shapes, y_n represents the original value of $(N * C * H * W)$, N represents the number of images, C represents the number of image channels, H represents the height of image, W the width.

In this article we use Rectified Linear Units (ReLUs [20]) to replace traditional Sigmoid and Tanh as for activation functions, the document [19] has certified that in the deep training of CNN, the use of ReLUs is many times faster than the use of Tanh.

4 Experimental Result and Performance Analysis

4.1 Experimental Database

This article selects the LIVE3D image database [21] Phase I and Phase II released by University of Texas at Austin, among them, Phase I database has 20 pairs of reference images, 365 pairs of distorted images totally, with each pair of the image containing left view and right view of symmetrical distortion. The distorted type includes JPEG compressed distortion, JP2K compressed distortion, white gaussian noise (WN), Gaussian Blur (BLUR) and fast fading (FF) distortion; Phase II database includes 8 pairs of reference images, the difference of Phase I and Phase II database is: Phase II distorted image library contains symmetrical distortion and asymmetrical distortion, and the left and right views are spliced together for storage. The distribution of LIVE3D image library can be seen in Tab. 1:

Table 1: The type and quantity of image distortion in database of phase I and phase II

Database	Reference image	Distorted image	JP2K	JPEG	WN	BLUR	FF
Phase I	20	365	80	80	80	45	80
Phase II	8	360	72	72	72	72	72

Select 80% of images from Phase I and Phase II distorted image libraries respectively as training set in the experiment of this article, 10% of verification set and 10% of test set, with no overlapping between test set and training set. The image block conversed by phase congruency as input in the course of training. The final prediction score of image quality is gained by obtain the average scores of all the subblocks of each image. For the convenience of processing data, this article normalize formula min-max for Difference Mean Opinion Score (DMOS) that the dataset presents, DMOS value is difference quality mean score of reference image and distorted image, the large the DMOS value, the large the degree of image distortion.

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{9}$$

Among them, X_{norm} is the data after normalization, X is original value, X_{min} and X_{max} are the minimum value and maximum value of original value.

4.2 Performance Index of Quality Evaluation

In order to measure the property of objective image quality evaluation method, the objective prediction score and objective evaluation of images need to be compared, at present days, the correlation of objective prediction value and objective evaluation are detected by Linear Pearson Correlation Coefficient (LPCC) and Spearman Rank Order Correlation Coefficient (SROCC). See Eqs. (10) and (11):

$$LPCC = \frac{\sum_{i=1}^{N} \left(X_i - \overline{X} \right) * \left(Y_i - \overline{Y} \right)}{\sqrt{\sum_{i=1}^{N} \left(X_i - \overline{X} \right)^2} * \sqrt{\sum_{i=1}^{N} \left(Y_i - \overline{Y} \right)^2}} \tag{10}$$

In (10): X_i and Y_i represent objective evaluation value and subjective prediction value of I image respectively, $\overline{X}$ and $\overline{Y}$ mean value of objective evaluation value and subjective prediction value, N represents the number of images. LPCC evaluate the precision of prediction, that is degree of accuracy, value range is within [0, 1], the closer to 1, the higher the degree of accuracy is, vice versa, the lower.

$$SROCC = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N \left(N^2 - 1 \right)} \tag{11}$$

In (11), N represents the number of images, d_i represents the difference between the objective evaluation value and subjective prediction value of image. SROCC evaluate the monotonicity of prediction, the value range is within [−1, 1], The absolute value is close to 1, the better the monotonicity.

4.3 Experimental Result and Analysis

To verify the effectiveness of the methods in the article, compare the experimental results on the LIVE3D, relatively new methods of FR image quality evaluation method with NR ones, the Tabs. 2–5

represent result comparison of different methods FR and NR stereo picture quality evaluation on the Phase I and Phase II image library. The models in this article are mainly targeting at JPEG, JP2K, WN, Blur, FF and various distorted types, we can see the methods in the articles are better than other stereo image quality evaluation model from the Tabs. 2–5. We can find from the documents [4–7] and the documents [9–12] that these evaluation models are mostly aiming at distortion of single type, having bad effect on the 5 distorted types combined. In the real life, the distorted image cannot always be pointed out clearly which type of distortion can be, this article, through carry outing comprehensive training on these 5 distorted types of JPEG, JP2K, WN, Blur, FF, no need to extract the inspection figure, structure features, high-frequency energy, and others of stereo image, we can see LPCC is obviously excel other evaluation model overall. But the models have a better evaluation effectiveness of JP2K, WN, Blur, FF of distorted types than JPEG distorted types, excel other stereo image quality evaluation models by and large.

Table 2: Comparison of LPCC results of different stereo image quality evaluation algorithms in phase I image database

Model method		JP2K	JPEG	WN	Blur	FF	ALL
Document [4]	FR	0.942	0.659	0.945	0.948	0.804	0.919
Document [5]	FR	0.877	0.487	0.941	0.919	0.730	0.930
Document [6]	FR	0.916	0.634	0.943	0.941	0.785	0.916
Document [7]	FR	0.923	0.656	0.941	0.951	0.840	0.924
Document [9]	NR	0.977	0.877	0.974	0.973	0.920	0.950
Document [10]	NR	0.951	0.811	0.963	0.961	0.870	0.895
Document [11]	NR	0.941	0.576	0.939	0.950	0.736	0.926
Document [12]	NR	0.979	**0.915**	0.950	0.954	0.912	0.944
Model of this paper	NR	**0.982**	0.720	**0.976**	0.960	**0.975**	**0.949**

Table 3: Comparison of SROCC results of different stereo image quality evaluation algorithms in phase I image database

Model method		JP2K	JPEG	WN	Blur	FF	All
Document [4]	FR	0.921	0.666	**0.948**	0.928	0.779	0.923
Document [5]	FR	0.860	0.439	0.940	0.882	0.588	0.924
Document [6]	FR	0.895	0.558	0.948	0.926	0.687	0.915
Document [7]	FR	0.875	0.614	0.943	0.937	0.781	0.921
Document [9]	NR	0.957	**0.950**	0.963	0.911	0.890	0.961
Document [10]	NR	0.909	0.767	0.921	0.867	0.820	0.891
Document [11]	NR	0.886	0.555	0.914	0.880	0.667	0.912
Document [12]	NR	0.950	0.867	0.929	**0.969**	0.887	**0.937**
Model of this paper	NR	**0.951**	0.738	0.800	0.933	**0.900**	0.892

Table 4: Comparison of LPCC results of different stereo image quality evaluation algorithms in phase II image database

Model method		JP2K	JPEG	WN	Blur	FF	ALL
Document [4]	FR	0.887	0.405	0.961	0.942	0.861	0.731
Document [5]	FR	0.905	0.830	0.912	0.784	0.915	0.774
Document [6]	FR	0.842	0.842	0.960	0.965	0.909	0.906
Document [7]	FR	0.837	0.750	0.849	0.827	0.880	0.758
Document [9]	NR	0.951	0.955	0.960	0.980	0.963	0.900
Document [10]	NR	0.875	0.796	0.956	0.949	0.911	0.880
Document [11]	NR	0.918	0.717	0.950	0.976	0.938	0.905
Document [12]	NR	0.901	**0.964**	0.914	0.934	0.901	0.915
Model of this paper	NR	**0.933**	0.847	**0.969**	**0.982**	**0.947**	**0.943**

Table 5: Comparison of SROCC results of different stereo image quality evaluation algorithms in phase II image database

Model method		JP2K	JPEG	WN	Blur	FF	ALL
Document [4]	FR	0.886	0.501	**0.956**	0.900	0.859	0.690
Document [5]	FR	0.894	0.795	0.905	0.813	0.891	0.720
Document [6]	FR	0.833	0.839	0.955	0.909	0.889	0.901
Document [7]	FR	0.847	0.719	0.845	0.800	0.850	0.745
Document [9]	NR	0.970	0.898	0.983	0.837	0.901	0.890
Document [10]	NR	0.842	0.693	0.925	0.879	0.933	0.880
Document [11]	NR	0.913	0.731	0.931	0.919	**0.944**	0.890
Document [12]	NR	0.954	0.886	0.901	**0.920**	0.904	0.901
Model of this paper	NR	**0.954**	**0.894**	0.906	0.894	0.904	**0.937**

5 Conclusion

This article puts forward a NR stereo picture quality evaluation model based on transfer learning, integrate the left view with the right view through PCA, and applies training features of CNN and regression models among image DMOS values, which effectively increase the performance of image quality evaluation model. Different from traditional learning, this article trains the progression model of stereo picture by trained parameter in the flat image library through transfer learning, overcoming the disadvantages of extracting features by hand, poor generalization ability and others in traditional machine learning. The experimental result shows that the models in this article is superior to some existing stereo picture quality evaluation model, but they are not effective on the JPEG distortion type, need improving, additionally, as for the convolution neutral network parameter setting of in-depth learning need further study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] I. T. Union, "Methodology for the subjective assessment of the quality of television pictures," *ITU-R Recommendation BT*, vol. 1, pp. 1–48, 2002.

[2] Z. Wang, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[3] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, A. C. Bovik, Z. Wang *et al.,* "Image and video quality assessment research at live," Available: http://live.ece.utexas.edu/research/quality/.

[4] J. You, "Perceptual quality assessment for stereoscopicimages based on 2D image quality metrics and disparity analysis," *Int. Workshop on Video Processing & Quality Metrics for Consumer Electronics*, Scottsdale, AZ, USA, pp. 1–6, 2012.

[5] F. Shao, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society*, vol. 22, no. 5, pp. 1940–1953, 2013.

[6] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems & Signal Processing*, vol. 24, no. 2, pp. 281–316, 2013.

[7] J. Zhang and Q. B. Sang, "Quality assessment method of color stereoscopic images," *Journal of Computer Applications*, vol. 35, no. 3, pp. 816–820, 2015.

[8] Y. Fang, J. Yan and J. Wang, "No reference quality assessment for stereoscopic images by statistical features," in *Ninth Int. Conf. on Quality of Multimedia Experience*, IEEE, Athens, Greece, pp. 1–6, 2017.

[9] M. Ling and Y. U. Mei, "No-reference stereoscopic image quality assessment based on binocular fusion and binocular rivalry," *Journal of Ningbo University (Natural Science & Engineering Edition)*, vol. 47, pp. 55–62, 2016.

[10] M. Liu, "No-reference stereoscopic image quality assessment based on deep feature learning," Tianjin University. Diss. pp. 45–48, 2018.

[11] R. S. Xiong and C. F. Li, "No-reference stereoscopic image quality assessment based on wavelet transform," *Computer Science*, vol. 42, no. 9, pp. 282–284, 2015.

[12] W. J. Tian, "Blind image quality assessment for stereoscopic images via deep learning," *Journal of Computer _Aided Design & Computer Graphics*, vol. 28, no. 6, pp. 968–975, 2016.

[13] S. Z. Yang, "Sample and feature based transfer learning method and Its application," *Diss. National University of Defense Technology*, vol. 4, no. 1, pp. 26–28, 2016.

[14] A. Noor, "Hybrid image fusion method based on discrete wavelet transform (DWT), principal component analysis (PCA) and guided filter," in *2020 First Int. Conf. of Smart Systems and Emerging Technologies (SMARTTECH)*, pp. 138–143, 2020.

[15] M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern Recognition Letters*, vol. 6, pp. 303–313, 1987.

[16] C. F. Li, G. F. Tang and X. J. Wu, "No-reference image quality assessment with lerning phase congruency feature," *Journal of Electronics &Information Technology*, vol. 35, no. 2, pp. 484–488, 2013.

[17] D. Yang and S. Q. Yu, "Image quality assessment based on phase congruency," *Computer Engineering & Applications*, pp. 1–2, 2015.

[18] Y. Jia, "Caffe: Convolutional architecture for fast feature embedding," *ACM*, pp. 675–678, 2014.

[19] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 1097–1105, 2012.

[20] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *ICML*, vol. 16, pp. 807–814, 2010.

[21] M. J. Chen, "Full-reference quality assessment of stereoscopic images by modeling binocular rivalry," *Signals, Systems & Computers IEEE*, vol. 5, pp. 721–725, 2013.