Jagannath University Journal of Science
Volume 08, Number I, December 2021, pp. 13–22
https://jnu.ac.bd/journal/portal/archives/science.jsp
ISSN 2224-1698

JnUJSci

Comparison of Loss Functions in Blind Image Quality Assessment using a Deep Bi-linearly Pooled ConvNet

Research Article

Md Rafiqul Islam, Mehrab Hosen Mahi, Arnisha Akhter, Md. Abu Layek*

Department of Computer Science and Engineering, Jagannath University, Dhaka-1100, Bangladesh;

Received: 30 December 2020
Accepted: 08 September 2021

Abstract: Blind Image Quality Assessment (BIQA) is used to predict human perceptual image quality scores without actually knowing the reference images. State-of-the-art methods typically require human subjects to score a huge number of image data for a robust model at the time of the training phase. Moreover, subjective quality scores are discriminating, inexact, and irrational. It is very hard to obtain a large-scale database or to extend existing databases, because of the difficulty of collecting images, training the subjects, conducting subjective experiments, and reconstructing human-level quality evaluations. The proposed method can assess both synthetically and authentically distorted images by using a method called 'Deep Bi-linear Pooling'. Two separate convolutional neural network models are used. One is used for synthetically distorted images and another is used for authentically distorted images. For synthetic distortion, a pre-trained CNN is used to classify the distortion type with available ground truth labels. And for authentically distorted images, a pre-trained CNN (VGG-16) is used which is trained with A large-scale database called 'Image-Net'. Then two networks are merged with the help of a method called 'Bi-linear' pooling. After fine-tuning the whole model, we have got state-of-the-art results for both the synthetically and the authentically distorted IQA datasets. For pre-trained synthetic images, the Waterloo Exploration Database is used. Here in this paper, we have tried to identify which Loss function can be used robustly and gives the better result relative to the most used loss function such as L1 Loss or Mean Absolute Error, L2 loss or Mean Squared Error, Smooth L1 Loss or Huber Loss, Poisson Loss.

Keywords: *Deep Learning • CNN • Bi-linear Pooling • Blind Image Quality assessment • Loss Function • Smoothl1loss • Poisson loss*

1. Introduction

Currently, mobile phones and various stationary devices are used to capture digital images. These images are compacted by conventional and original approaches (Bovik, 2010; Lee *et al*., 2018) and are passed through different broadcasting mediums (Chen *et al*., 2019) as well as the images are saved in different storage forms. During the image capturing, proceeding, carrying, and warehousing, various unpredicted distortions can occur and in the meantime, perceptual information loss can happen with some image quality degradation. That's why Image Quality Assessment (IQA) becomes remarkably important for the image quality monitoring and reliability of image processing systems. The final judge of the perceptual image quality is the human visual system, so subjective IQA is the most well-grounded although it is time consuming and expensive. Because of this, objective

* Corresponding author: Md. Abu Layek
E-mail: layek@cse.jnu.ac.bd

IQA algorithms are developed for research labs to the actual real-world application (Maia *et al*., 2015)

Objective IQA can be categorized as three classes:

1) Full-reference IQA (FR-IQA)
2) Reduced-reference IQA (RR-IQA)
3) No-reference or Blind IQA (BIQA)

Now-a-days, BIQA is one of the most attractive research fields for researchers. Conventional BIQA models usually select low-level features- handmade (Mittal *et al*., 2012) or knowledgeable (Kang *et al*., 2014) to identify the degree of variations from analytical regularities of natural scenes, according to the quality prediction function is derived (Ma, 2017). For a long time, deep convolutional neural networks don't use because of lacking of adequate ground truth values such as MOS for training. MOS is expressed as Mean Opinion Scores. Some naive solutions are also used for quality prediction such as, fine-tune directly to pre-trained on Image-Net (Fei-Fei *et al*., 2009; Gao *et al*., 2018). These models acquire decent functioning on LIVE Challenge Database (Ghadiyaram and Bovik, 2015) (distortion types called authentically distorted images). However, the known and inherent characteristics of the IQA data sets, i.e. LIVE (Sheikh *et al*., 2006) and tid2013 (Ponomarenko *et al*., 2015) (distortion types called synthetically distorted images) prohibit them to achieve state-of-the-art performance for other types of distortions. On the other hand, image quality assessment, such as patch-based techniques gain quality score heirloom from the images or roughly by FR-IQA models (Kim and Lee, 2016). These models are effective for synthetically distorted images but not for authentically distorted images.

Throughout the solution of BIQA, not only the synthetically distorted images but also authentically distorted images are used. For synthetic distortions in the previous works (Ma *et al*., 2017; Yue *et al*., 2019; Moorthy and Bovik, 2010), a huge amount of pre-trained set constructed on 'Waterloo Exploration Data base' (Ma, Duanmu *et al*., 2016) and another large database called 'PASCAL VOC' (Everingham *et al*., 2010), where the synthetization process was done on images with 9 distortion levels.

When a multi-class CNN model is trained and the features are based on CNNs, the resulting architecture consists of standard CNN units for feature extraction, followed by a specially designed bilinear layer and a pooling layer. For example, the authentically distorted images are pre-trained using the ConvNet named VGG-16 (Simonyan and Zisserman, 2014) which is trained on Image-Net (Fei-Fei *et al*., 2009). These networks are then combined with a deep bilinear pooling (Ge *et al.*, 2019). By fine-tuning, the model learns on target data sets in form of the 'Stochastic Gradient Descent' or (SGD) methodology. The observation is that- Deep Bilinear Convolutional Neural Networks is more powerful than most of the today's ConvNet based network (Bosse *et al.*, 2017; Ma *et al.*, 2017).

Loss functions have the direct effect on the performance of any CNN network. With other factors, the loss functions also have application dependency. A loss function with a specific CNN giving better performance may not be the best suited for another application where another loss functions may perform better. The contribution of this paper lies in the investigation of loss functions for the bi-linearly pooled convolutional network where the target application is blind image quality assessment (BIQA).

1 Related Works

For the assessment of blind image quality, there are a few models used (Wang and Bovik, 2011; Ma, 2017; Ma and Liu *et al.*, 2017; Ye, 2014). A radial basis function was used by Tang et al. (Tang *et al.,* 2014) for the pre-trained belief net, and for predicting image quality it was fine-tuned. Biano et al. used different design options of ConvNet. For features classification they have used support vector regression (SVR). They have also pre-train the multi-class model by estimating MOS's into 5 classes. Despite all of these, their proposed models were not thoroughly optimized and demand huge physical parameter adaptations (Bianco *et al.,* 2018). Ye *et al.* used a ConvNet with huge counts of images for the quality scores by averaging predicted images by cropping from patches and normalizing spatially and calculated (Ye *et al.*, 2014). It is problematic in local perceptual quality when the quality measure of patches heirloom from equal images and also it is incompatible with global quality across spatial location; most of the time due to non-stationary of images content across spatial locations. Bosse *et al.* overcome these problems by considering two approaches (Bosse *et al.*, 2017):

1) To take the mean value of multiple patches.
2) To take the weight mean quality counts of patches by comparative significance.

Kim et al. and many other researchers have used deep CNN for image quality assessment (Kim and Lee, 2016). Table I is presents some previous works on image quality assessment (IQA) and their used loss function and SLCC

scores. Deep bi-linear ConvNet (Zhang *et al.*, 2018) was proposed by W. Zhang, K. Ma, J, Yan, D. Deng, and Z. Wang where they used two networks for two different types of distorted images. One is called authentically distorted images and another is synthetically distorted images. Then the two networks are merged into one by a method called bi-pooling.

1.1 Loss function

Now-a-days, neural networks are very popular for computer vision and image processing and various architectures are developed to solve numerous problems in this field. The loss function is one of the crucial factors in deep learning. But most of the time, it doesn't get much interest to be focused. Actually, the l2 norm is chosen by default and virtually most of the time. It is shown that without changing the network model, and only by changing or taking different loss functions, the result could be more remarkable.

Table 1. Review of some previous works in turn of Loss function and SRCC score

Ref.	Loss function	Database	SRCC (%)
(Bosse *et al.*, 2017)	MAE(L1)	LIVE	0.96
(Bosse *et al.*, 2017)	MAE(L1)	TID2013	0.84
(Ma *et al.*, 2017)	Empirical Cross Entropy Loss	TID2013	0.81
(Ma and Liu *et al.*, 2017)	RankNet's Loss Function	LIVE	0.96
(Ma and Liu *et al.*, 2017)	RankNet's Loss Function	TID2013	0.90

1.2 Some well-known loss functions

A. Mean Absolute Error (L1) Loss: The Mean Absolute Error loss is a good loss function where it is more remarkable to outliers. It is calculated by averaging of the absolute difference between the actual and predicted values. The formula of MAE is given in the equation 1.

$$loss(x, y) = \sum_{i=1}^{N} ||x_i - y_i|| \quad (1)$$

Graphical representation of the loss is shown in Figure 1.

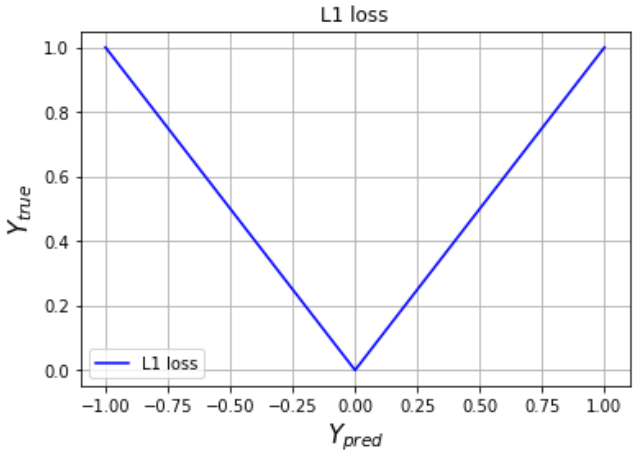

Figure 1. Mean Absolute Error (L1) Loss

A model is called a better model if it has a lower MAE value. The zero value is not acceptable because that has no significance and that's some sort of resource lost.

Mean Squared Error (L2) loss:

The Mean Squared Error loss is mainly used when the problem is based on regression. In mathematics, if the target variable is distributed in Gaussian, then l2 loss is preferred. If there are no other remarkable issue, it is the most used loss function. The formula for MSE is given in the equation 2.

$$loss(x, y) = \sum_{i=1}^{N} (x_i - y_i)^2 \quad (2)$$

Graphical representation of the loss is shown in Figure 2.

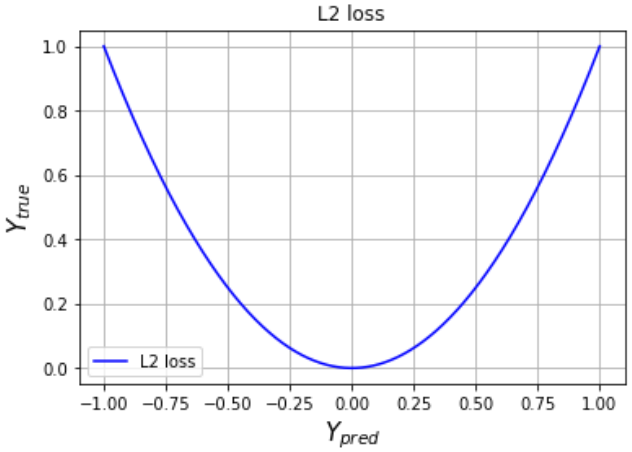

Figure 2. Mean Squared Error (L2) loss

Average of the squared differences between the predicted and actual values is used for calculating the loss. The value of the loss function is always positive.

There are some works in the context of computer vision tasks (Stadelmann *et al.*, 2019). Neural networks that are used for denoising (Gondara, 2016; Xie *et al.*, 2012; Tao *et al.*, 2018), all use the l2 norm.

L2 norm is arguably the dominant normalization function for a long time because it is convex and differentiable and has very convenient properties for optimization problems, pattern recognition, signal processing, image processing etc. L2 provides the highest probability estimates in the case of liberated and similar distributed Gaussian noise.

There are some more loss functions, such as Poisson loss, SmoothL1Loss or Huber loss, Kullback-Leibler divergence, Cross-Entropy Loss etc. None of them is used widely, rather problem specific.

2 Experimental Result

2.1 Experiment on LIVE data-set

A. L1 loss Function: From the Figure 3, we can see that over the 10 session of the LIVE data set, sessions 4 and 5 and 8 are well forms. Others are not so suited. Data are skewed. Test data assessment scores have never reached the training data assessment scores.

B. L2 loss Function: From the Figure 4, we can see that over 10 sessions on the LIVE data set have used the L2 loss function, test date score has never given better value. Most probably, the set is not suitable for the L2 loss function or the dataset is overfitted. In the original paper (Zhang *et al.*, 2018), they have used the L2 loss function. But in our experiment, the L2 loss function does not yield the overall best result. The test assessment score and train assessment are score is very different in a large number of values.

C. SmoothL1Loss Function: From the Figure 5, we can see that this is far better than Fig: 4 in sense of all sessions. But still, it is not very good loss function for the LIVE dataset. In session 1, it gives the best result, but in sessions 8 and 10, result changes are noticeable. Test scores vary drastically throughout the whole session.

D. Poisson Loss Function: From the Fig 6, we can again see that the Poisson loss function is good for the LIVE data set. In every session, the test score and train score are much different in a large number of value. Not a single session has given a satisfactory result. Almost in every session, the test score is so little and changes drastically.

2.2 Experiment on TID-2013 data-set

A. L1 loss Function: From Figure 7, we can analyze that TID2013 is working fine in every session. In sessions 1, 4, 7, 8 the training scores and testing scores are very close to each other. In sessions 2 and 5, testing scores are always higher than training scores. We can say that it could have happened that in this session the model is under-fitted.

B. L2 loss Function: From Figure 8, we can see that sessions 2, 4, 7, 8, and 10 are working very well. But in session 1, the testing score never rise after the 25th epochs even though training scores rose after the 40th epoch. In sessions 5 and 10, the testing and training scores rise at the close to the 50th epoch. So we can predict that accuracy could be much better if we use more epochs. In sessions 6 and 9, the testing scores are downward even though the training scores are rising. It may be the cause of overfitting.

C. SmoothL1Loss Function: From Figure 9, it is shown that in every session the model is over-fitted in some sessions and under fitted in some others. In sessions 1, 3, 6, and 9, the model is over-fitted and in sessions 2, 5, and 8 the model is under-fitted. Sessions 3, 5, 6, and 8 have given relatively better results in training and testing scores. The differences between the training and testing are very low which is actually good.

D. Poisson Loss Function: From Figure 10, for Poisson loss function, the result is not good. Almost every session, the model under-fitted. In sessions 1, 2, 3, 4, 5, 7, 8, and 10 are very much of the testing score than the training score. In session 2, the model is not learning much after the 40th epoch. In session 5 the model testing score always higher than the training score. The variation of the learning curve for training is affordable but the testing curve is not good because it is changing very often and the learning score is up and down.

3 Dataset Compare with Various Loss

3.1 LIVE Dataset

The LIVE dataset is a relatively small dataset. It contains around 779 images for training and testing. So, sometimes the model can be over-fitted.

It is often happening for small datasets that the training accuracy and the testing accuracy differ very much. It is also observed for the LIVE dataset too. Here in the Fig 11, it is shown. We have summarized all used loss functions in the table IV-A.

3.2 TID2013 Dataset

TID2013 is a relatively larger dataset. It contains around 3000 images for training and testing purposes. So we can say it is a pretty good dataset for working with large models.

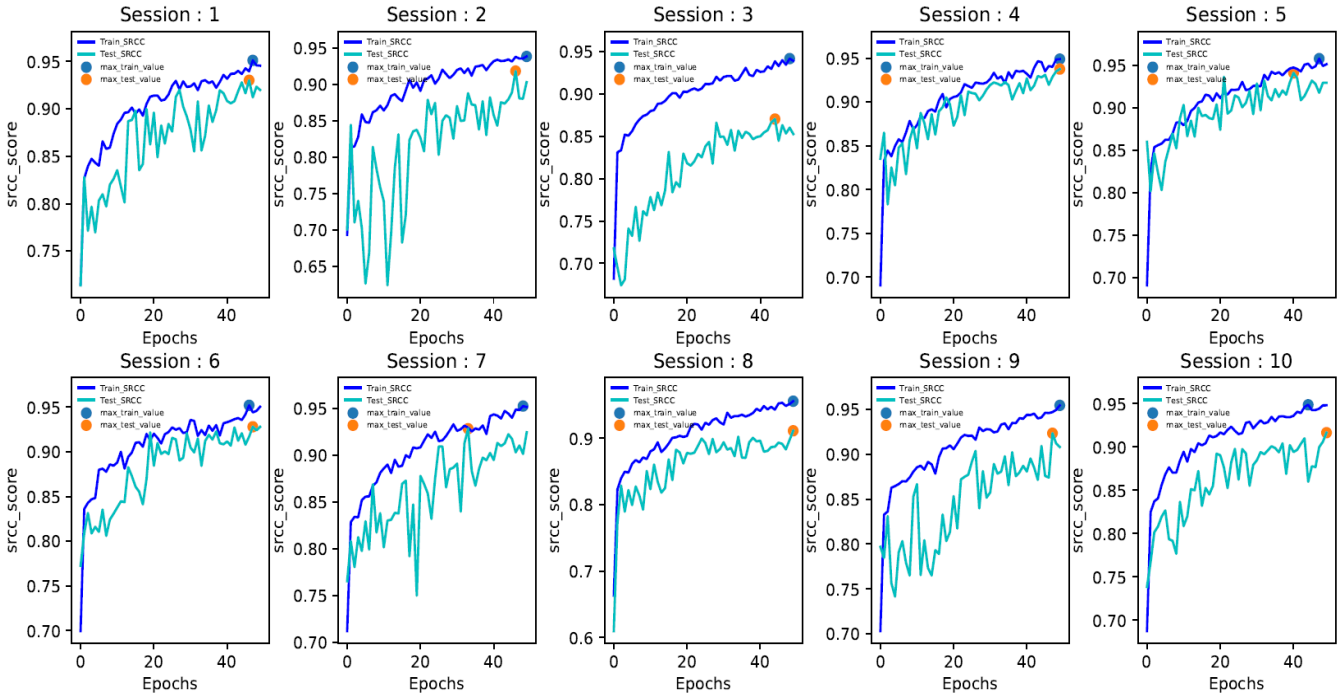

Figure 3. Graphical presentation of 10 session of L1 Loss function on LIVE

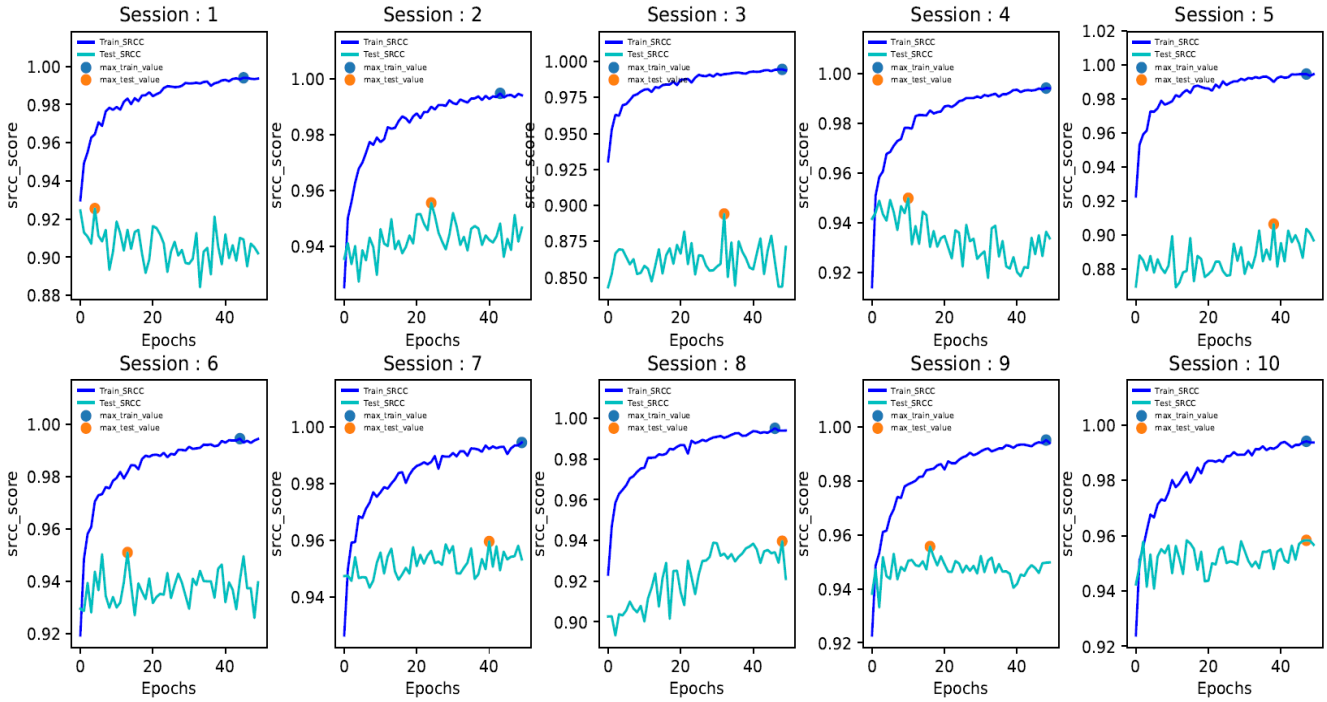

Figure 4. Graphical presentation of 10 sessions of L2 Loss function on LIVE

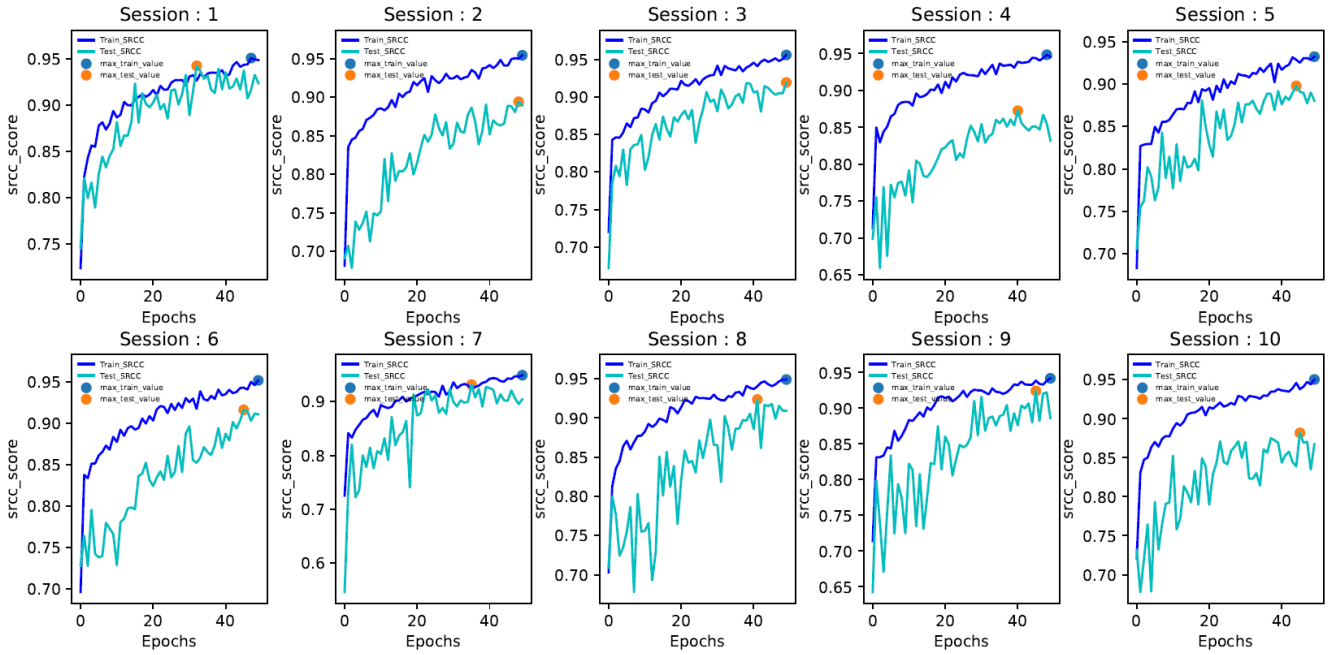

Figure 5. Graphical presentation of 10 sessions of SmoothL1Loss function on LIVE

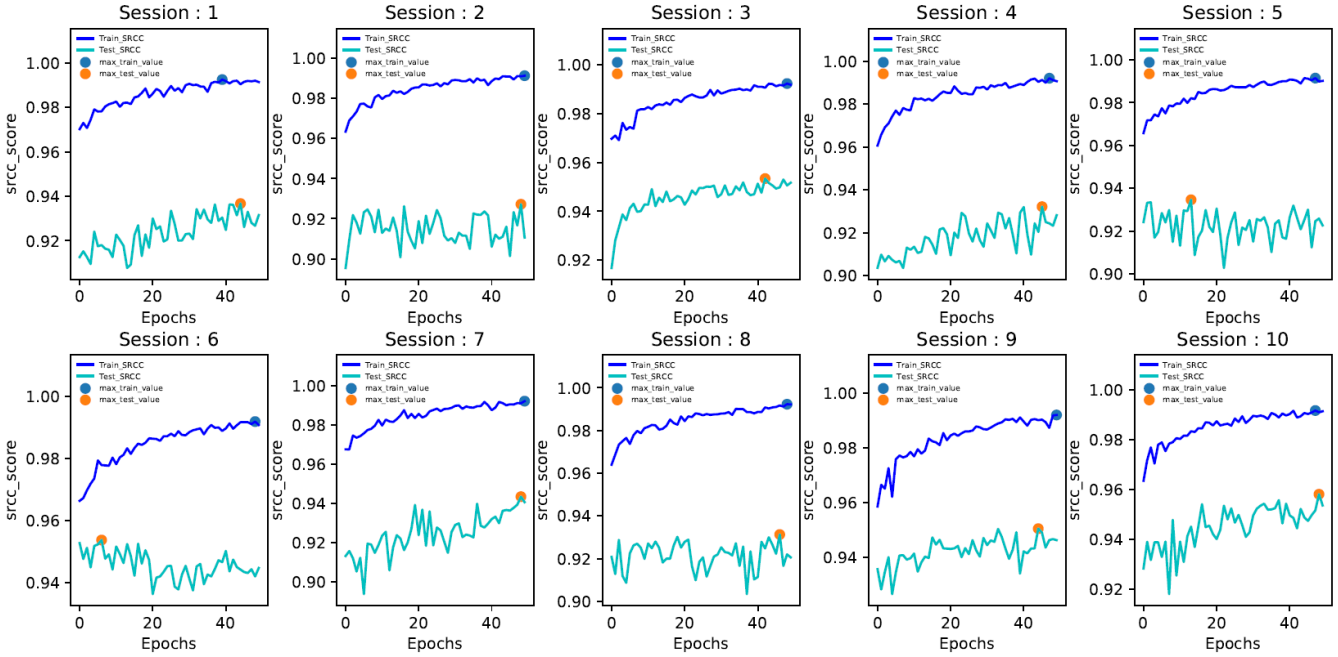

Figure 6. Graphical presentation of 10 sessions of Poisson Loss function on LIVE

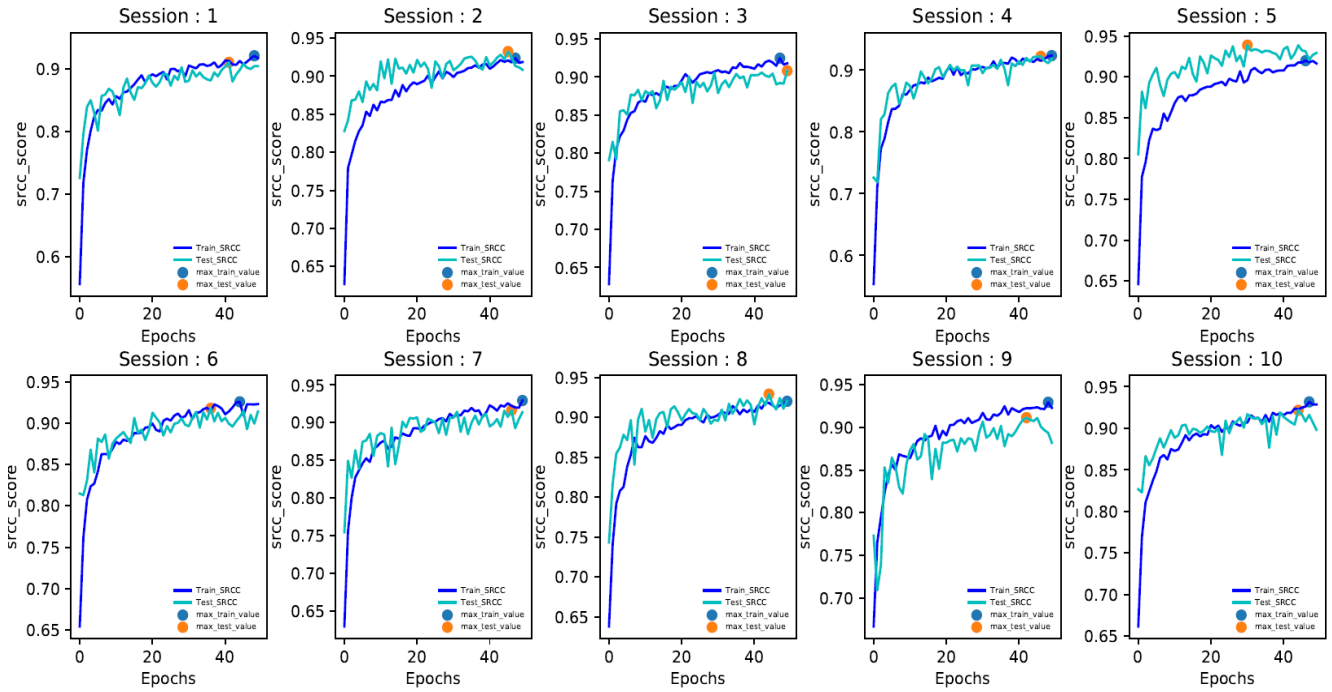

Figure 7. Graphical presentation of 10 sessions of L1 Loss function on TID2013

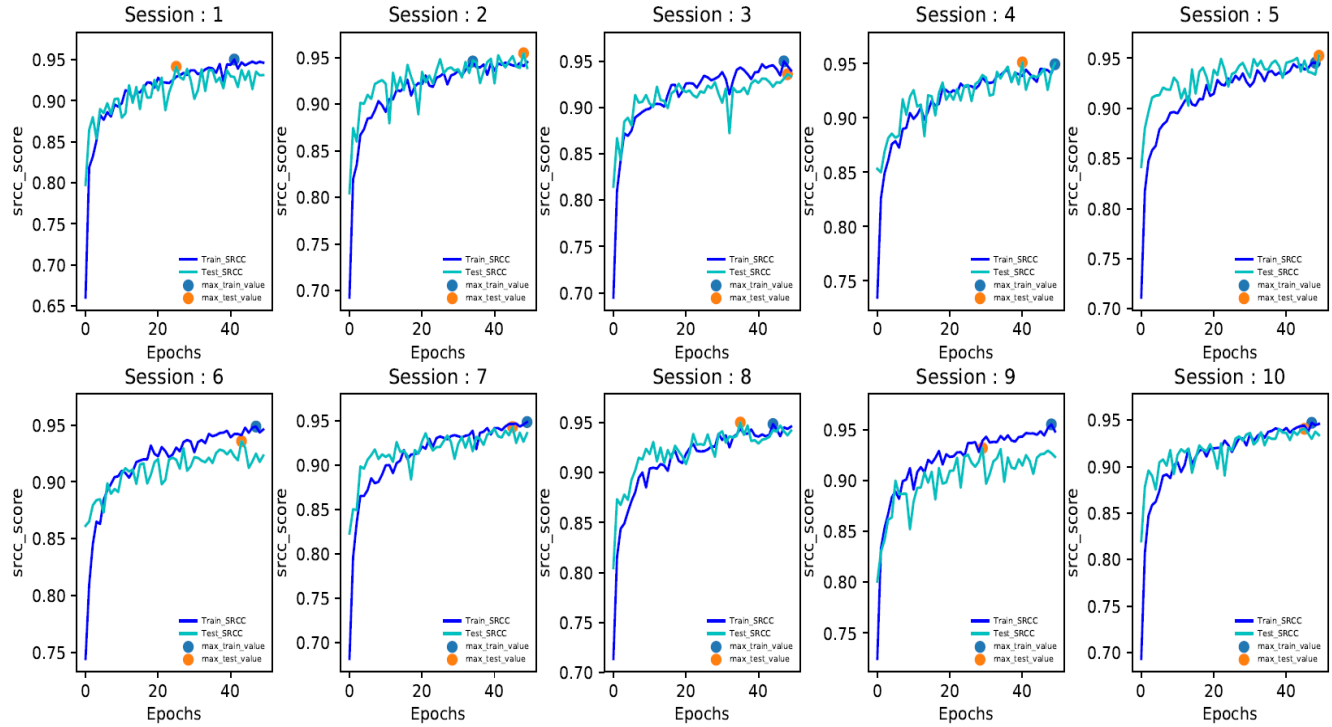

Figure 8. Graphical presentation of 10 session of L2 Loss function on TID2013

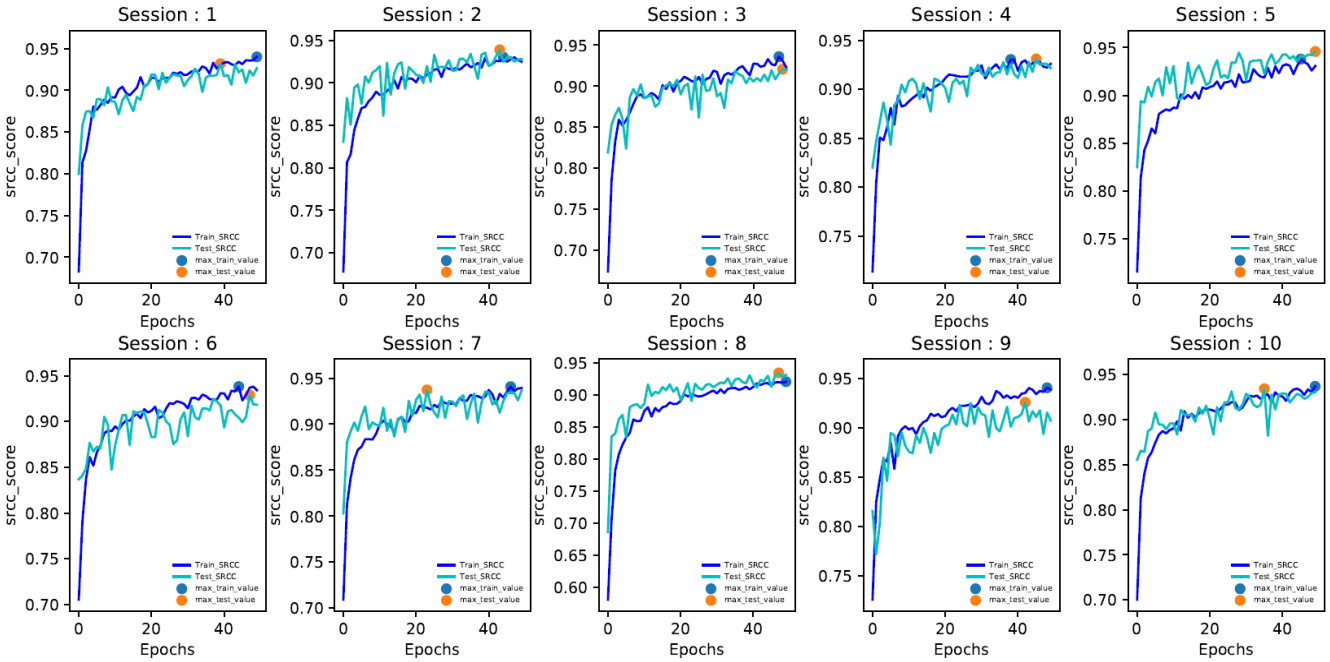

Figure 9. Graphical presentation of 10 sessions of **Smoothl1loss** function on TID2013

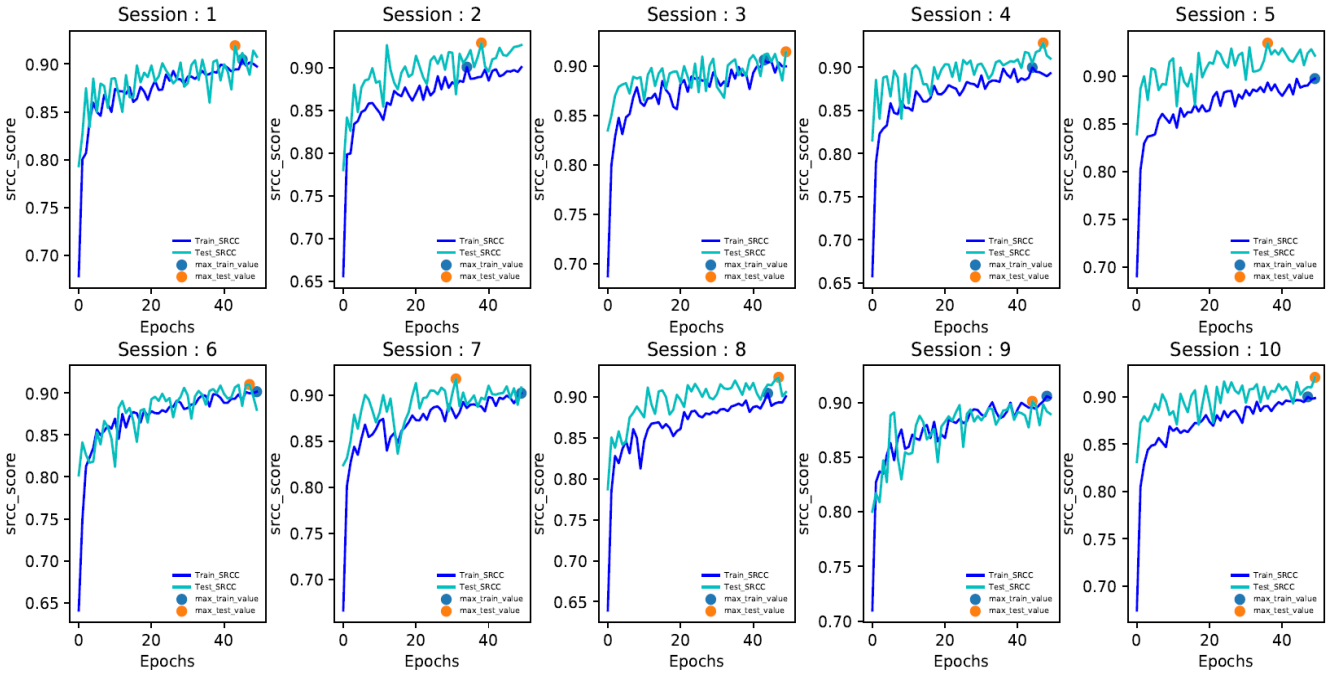

Figure 10. Graphical presentation of 10 sessions of Poisson Loss function on TID2013

For TID2013, the model works fine for the training and testing phase. The difference between the training score and testing is so low; sometimes almost equal to some fractions. Every loss function used is giving the same result but the score varies. For L2, it gives the best result. Although, it is not very high relatively SmoothL1 Loss. For Poisson loss, the training score is lower than testing. So we can say, for this loss function, the TID2013 data set is not good as it is under-fitted. It is represented in Fig 12. We have summarized the all used loss functions as shown in table IV-B.

4 Conclusions

We have tried to identify the relationship of various loss functions along with deep bi-linear convolutional

neural networks with different available data sets. The result is impressive that the various loss functions yield. Different data sets also show different accuracy when the same network is used, such as L1, L2, SmoothL1Loss, Poisson Loss, etc. When data sets are distributed in poison distribution, then the Poisson loss is effective.

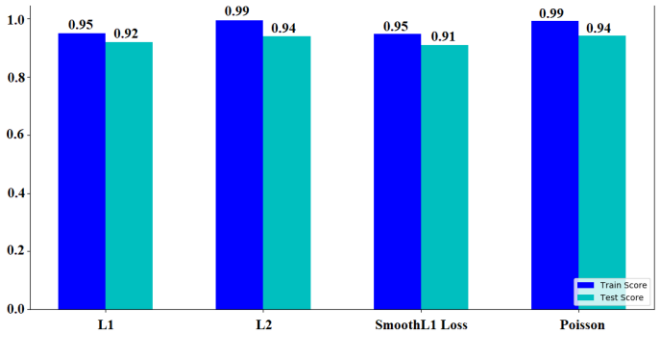

Figure 11. Summary of used loss functions on LIVE

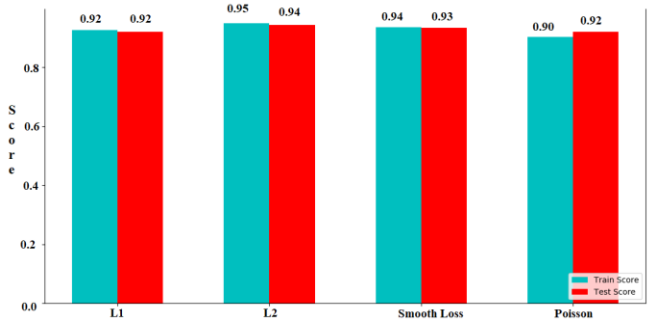

Figure 12. Summary of used loss functions on TID2013

Table 2. Experimental result on live dataset

Loss Function Name	Training Avg(%)	Testing Avg(%)
L1 loss or MAE	95%	92%
L2 loss or MSE	99%	94%
SmoothL1Loss	95%	91%
PoissonLoss	99%	94%

In the most cases where the problem is regression-based, L2 loss or MSE loss is used without any other consideration. But we have to observe the data set before applying the MSE loss function. It could be useful to use other loss functions too.

Table 3. Experimental result on tid-2013 dataset

Loss Function Name	Training Avg(%)	Testing Avg(%)
L1 loss or MAE	92%	92%
L2 loss or MSE	95%	94%
SmoothL1Loss	94%	93%
PoissonLoss	90%	92%

Acknowledgements

This work was supported by the Information Technology Research and Resource Center (ITRRC, web: http://itrrc.com), and the JnU research grant (জবি/গবেষণা/গপ্র/২০২০-২০২১/বিজ্ঞান/৩৩) Jagannath University, Dhaka, Bangladesh.

References

Bianco, S., Celona, L., Napoletano, P., & Schettini, R. (2018). On the use of deep learning for blind image quality assessment, *Signal, Image and Video Processing*, 12(2) : 355–362.

Bosse, S., Maniry, D., Muller, K.-R., Wiegand, T., & Samek, W. (2017). Deep neural networks for no-reference and full-reference image quality assessment, *IEEE Transactions on Image Processing*, 27(1): 206–219.

Bovik, A. C. Handbook of image and video processing. Academic press, 2010.

Chen, P., Li, L., Huang, Y., Tan, F., & Chen, W. (2019). QoE evaluation for live broadcasting video, *In 2019 IEEE International Conference on Image Processing (ICIP), IEEE*, 454-458.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisser-man, A. (2010). The pascal visual object classes (voc) challenge, *International journal of computer vision*, 88 (2): 303–338.

Fei-Fei, L., Deng, J., & Li, K. (2009). ImageNet: Constructing a large-scale image database, *Journal of vision*, 9(8): 1037-1037.

Gao, F., Yu, J., Zhu, S., Huang, Q., & Tian, Q. (2018). Blind image quality prediction by exploiting multi-level deep representations, *Pattern Recognition*, 81: 432-442.

Ge, S. Y., Gao, Z. L., Zhang, B. B., & Li, P. H. (2019). Kernelized Bilinear CNN Models for Fine-Grained Visual Recognition, *ACTA ELECTONICA SINICA*, 47(10): 2134.

Ghadiyaram, D., & Bovik, A. C. (2015). Massive online crowdsourced study of subjective and objective picture quality, *IEEE Transactions on Image Processing*, 25(1): 372–387.

Gondara, L. (2016). Medical image denoising using convolutional denoising autoencoders, *In 2016 IEEE 16th international conference on data mining workshops (ICDMW), IEEE*, 241-246.

Jain, V., & Seung, S. (2009). Natural image denoising with convolutional networks, *Advances in neural information processing systems*, 769–776.

Kang, L., Ye, P., Li, Y., & Doermann, D. (2014). Convolutional neural networks for no-reference image quality assessment, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1733-1740.

Kim, J., & Lee, S. (2016). Fully deep blind image quality predictor, *IEEE Journal of selected topics in signal processing*, 11(1): 206– 220.

Lee, J., Cho, S., & Beack, S. K. (2018). Context-adaptive entropy model for end-to-end optimized image compression, *arXiv preprint arXiv*, 1809. 10452.

Ma, K. Blind image quality assessment: Exploiting new evaluation and design methodologies, 2017.

Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., & Zhang, L., (2016). Waterloo exploration database: New challenges for image quality assessment models, *IEEE Transactions on Image Processing*, 26(2): 1004–1016.

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

Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., & Zuo, W. (2017). End-to-end blind image quality assessment using deep neural networks, *IEEE Transactions on Image Processing*, 27(3): 1202–1213.

Maia, O. B., Yehia, H. C., & de Errico, L. (2015). A concise review of the quality of experience assessment for video streaming, *Computer communications*, 57: 1-12.

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain, *IEEE Transactions on image processing*, 21(12): 4695–4708.

Moorthy, A. K., & Bovik, A. C. (2010). A two-step framework for constructing blind image quality indices, *IEEE Signal processing letters*, 17(5): 513–516.

Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., (2015). Image database tid2013: Peculiarities, results and perspectives, *Signal processing: Image communication*, 30: 57–77.

Sheikh, H. R., Sabir, M. F., & Bovik, A. C. (2006). A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on image processing*, 15(11): 3440–3451.

Simonyan, K., & Zisserman, A., (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv*, 1409.1556.

Stadelmann, T., Tolkachev, V., Sick, B., Stampfli, J., & Dürr, O. (2019). Beyond ImageNet: deep learning in industrial practice, *In Applied Data Science*, *Springer, Cham*, 205-232.

Tang, H., Joshi, N., & Kapoor, A. (2014). Blind image quality assessment using semi-supervised rectifier networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2877–2884.

Tao, X., Gao, H., Shen, X., Wang, J., & Jia, J. (2018). Scale-recurrent network for deep image deblurring, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8174-8182.

Wang, Z., & Bovik, A. C. (2011). Reduced-and no-reference image quality assessment, *IEEE Signal Processing Magazine*, 28(6): 29– 40.

Xie, J., Xu, L., & Chen, E. (2012). Image denoising and inpainting with deep neural networks, *In Advances in neural information processing systems*, 341-349.

Ye, P. Feature learning and active learning for image quality assess-ment, Ph.D. dissertation, 2014.

Yue, G., Hou, C., Yan, W., Choi, L. K., Zhou, T., & Hou, Y. (2019). Blind quality assessment for screen content images via convolutional neural network, *Digital Signal Processing*, 91, 21-30.

Zhang, W., Ma, K., Yan, J., Deng, D., & Wang, Z. (2018). Blind image quality assessment using a deep bilinear convolutional neural network, *IEEE Transactions on Circuits and Systems for Video Technology*.