Deeplearning Convolutional Neural Network based QoE Assessment Module for 4K UHD Video Streaming

Akm Ashiquzzaman[1][a], Sung Min Oh[1], Dongsu Lee[1], Hoehyeong Jung[1], Tai-won Um[2] and Jinsul Kim[1]

[1]*School of Electronics and Computer Engineering, Chonnam National University, Gwangju, South Korea*

[2]*Department of Information and Communication Engineering, Chosun University, Gwangju, South Korea*

Keywords: Deep Learning, Convolutional Neural Network, Computer Networks, Video Steaming, 4K UHD, QoE.

Abstract: With the rapid development of modern high resolution video streaming services, providing high Quality of Experience (QoE) has become a crucial service for any media streaming platforms. Most often it is necessary of provide the QoE with NR-IQA, which is a daunting task for any present network system for it's huge computational overloads and often inaccurate results. So in this research paper a new type of this NR-IQA was proposed that resolves these issues. In this work we have described a deep-learning based Convolutional Neural Network (CNN) to accurately predict image quality without a reference image. This model processes the RAW RGB pixel images as input, the CNN works in the spatial domain without using any hand-crafted or derived features that are employed by most previous methods. The proposed CNN is utilized to classify all images in a MOS category. This approach achieves state of the art performance on the KoniQ-10k dataset and shows excellent generalization ability in classifying proper images into proper category. Detailed processing on images with data augmentation revealed the high quality estimation and classifying ability of our CNN, which is a novel system by far in these field.

1 INTRODUCTION

In the edge of 5G telecommunication and high speed 4K UHD streaming technologies, mandatory delivery of high quality content has become a crucial task. by 2025, the 4K UHD streaming will become the main standard of video streaming technologies. Universal deployment of Dynamic Adaptive Streaming of HTTP (DASH) is also becoming popular for this streaming technologies (Van Ma et al., 2018). Popular video streaming services such as Youtube, Netflix, Amazon Hulu, etc. are moving to these standard recently. These streaming technologies are mainly focusing on full service and experience provided to the users to make sure they have the maximum satisfaction in service and quality.

In the past, the concept of Quality of Service (QoS) has been brought to the attention of a large number of users as well as service providers and network operators. QoS also has high the investment rate of telecom operators as well as high concentration of the network research community, leading to solutions that ensure highly stable and highly efficient (Yoo et al., 2003). Compared with the concept of QoS, QoE is a newer concept. QoS simply gives users a fairly technical sense of service quality. QoS is primarily focused on describing the objective, technical criteria that the network infrastructure or application needs to achieve in order to guarantee quality service. In other words, QoS can be considered as the general technical language of the quality that applications and network infrastructure used. Consequently, we must establish a general description model which easy to understand for end-users of a service. That is the concept of QoE. In fact, QoE is a common language for applications and end users (Nam et al., 2016). It is used in the approach of a user to evaluate QoS. In other words, QoE is a measure of the satisfaction of people with the service they are using based on subjective judgment. Thus, QoE can be synthesized from pure QoS and other non-technical factors such as the characteristics of the human visual and auditory system, etc. The main differences can be distinguished in the Figure 1. The emergence of the QoE concept will most likely lead to certain changes in the market approach of video streaming service providers. Rather

[a] https://orcid.org/0000-0002-2215-8576

Ashiquzzaman, A., Oh, S., Lee, D., Jung, H., Um, T. and Kim, J.
Deeplearning Convolutional Neural Network based QoE Assessment Module for 4K UHD Video Streaming.
DOI: 10.5220/0008117903920397
In *Proceedings of the 9th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2019)*, pages 392-397
ISBN: 978-989-758-381-0

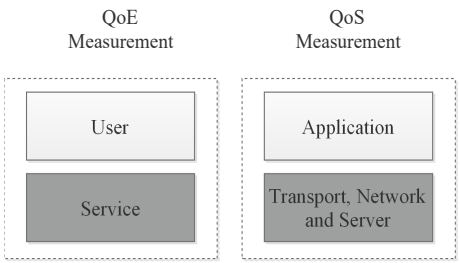

Figure 1: Comparison Between QoS and QoE shown in a simplified graph.

than focus on QoS, QoE-related issues will be centered.

In this Research, the main QoE assessment model was developed with the NF-IQA that is developed with CNN with classify the given images in it's distinctive MOS. The main idea of this CNN is to classify images that is being sampled in various time domain in a steaming in edge side then classify the MOS to asses the maximum MOS to ensure best QoE. The train CNN has outperformed the state of the art CNN models and later the CNN will deployed into the QoE assessment tool shown excellent result, that paves the way to deeplearning based CNN in the QoE assessment tools.

The rest of the paper is organized as follows. The Section 2 describes the previous works and Section 3 describes the dataset and the processing of it to train the CNN, Section 4 discusses the proposed methods of the model. Section 5 describes the experiment and result followed by the Conclusion in Section 6.

2 RELATED WORKS

To the best of our knowledge, deep learning CNN has not been applied to general-purpose QoE assessment Models. The primary reason is that the original CNN is not designed for capturing image quality features. In the object recognition domain good features generally encode local invariant parts, however, for the NR-IQA task, good features should be able to capture the aesthetics of the images as a whole. because of this problem, the CNN based QoE model has not yet been properly researched.

For providing the best QoE, the component has to master the Video Quality Assessment (VQA). In general, most of the module has to understand Image Quality Assessment(IQA), as video sampled in random time to ensure proper quality is essentially images. Visual quality is a very tiresome yet important property of an image. In principle, it is the calculated of the distortion compared with an ideal imaging model or perfect reference image. This type of system is usually know as the Full Reference (FR) IQA model (Sheikh et al., 2005). State of the art FR IQA models have always the best performance as it directly quantify the differences between distorted images and their corresponding ideal versions. As a result, this achieves a very high accuracy of correlated to human eyes, which is essentially the experience center.

In the technical sense, all the QoE model build based on this principles have to be ensured to provide the base or main reference image to compare. But the main drawback of such models are often the model has no reference image due to the configuration of the network of systems. This is also known as the NR (No Reference) IDA. As NR-IQA can directly asses image quality by exploiting features, it is easier to deploy in a standalone systems. So, the NR IQA gives the image quality by justifying the image aesthetics or the characteristics that is correlates to human eyes. Recently, deep neural networks have research and deemed optimal for recognitions and achieved great success on various computer vision tasks. Specifically, CNN (Convolutional Neural Network) has shown superior performance on many standard object recognition benchmarks (Krizhevsky et al., 2012). The main advantages of these models that it takes raw images, as in pixels in the input and incorporate feature learning to classify output. With a deep structure, the CNN can effectively learn complicated mappings while requiring minimal domain knowledge.

Kang et al. (Kang et al., 2014) described a CNN for no-reference image quality assessment. But the CNN take input as the grayscale rather than the RGB and had linear optimization process. This type of process is computationally enriched and often had problems in QoE model implementation.

This research was inspired by our previous research about MEC (Van Ma et al., 2018) with content-awareness component which is placed at MEC to retrieve DASH information for clients. On the basic of research on fuzzy logic to obtain DASH segments with high quality, we deploy segment selection for DASH streaming to MEC. As a result, it reduces network latency as well as the computation resource of clients with high streaming quality. However, the assessment module needs to be in a state of art NF IQA based model that can classify the hi resolution images and later it will adjust the service based on the quality. The CNN based research in this field is novel.

3 DATASETS

Neural network based research needs a high amount of reliable data for accurate learning. There should be a variety of data distributed in the label or class properly for the neural network to learn the distinguished features. A variety of IQA databases have been released nowadays for focusing mainly on the development. but most of the datasets are not focused on the quantity. As the main focus of the research was to develop the neural network or CNN for the VQA, the KonIQ-10k Dataset was the perfect choice for this research.

3.1 KonIQ-10k Dataset

In our research the main focus to development a fully automated QoE assesment model for the NF VQA. The model was trained eith the KonIQ-10k Dataset(Lin et al., 2018) for this purpose. in this dataset, 10,073 images were sampled from around 4.8 million images with 120 quality ratings, which were obtained the Mean Opnion Score (MOS) by crowd sourcing performed by a total of 1,467 crowd workers.

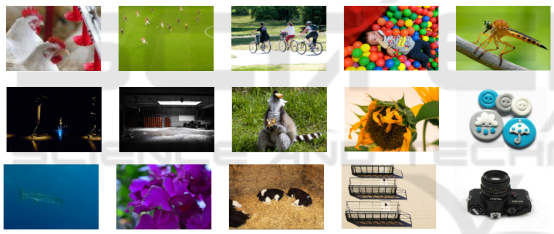

Figure 2: KonIQ-10k Dataset Image examples.

In Figure 2, some example images were seen that has variable MOS based on the aesthetic quality of the given images.

The main goal of deeplearning to generalized learning. In our study, we've focused into the image augmentation as our data mainly hovers around visual elements of images to illustrate various images states, basically augmenting the images makes it even more generalized for the neutal network to learn as it will not learn any focused area in single instances. This Augmentation expands the volume of training data, resulting more accuracy in generalized validation.

All the images in the dataset then later processed and classify into 4 categories of MOS scores based on the flooring of the original MOS scores and thus the values were transfers from continuous to a categorical ones. These images then used in data augmentation mode makes the train more robust and higher accuracy in classifying in model validation.

4 PROPOSED METHODS

4.1 Convolutional Neural Network

Hubel and Wiesel in 1960s demonastrated animal visual cortex contains neurons that perticullarly responds to specific area of the perspective field(Jung et al., 1963). So the theory of Convolutional Neural Network evolved by compute convolution in the input matrix to exploit edge features for the network to clssify and respond. Images, signal waves, sound, etc. are essentially digital signals that are represented in a multi-dimensional arrays. Lecunn et al. (LeCun and Bengio, 1995) first uses convolutional neural networks successfully with backpropagation algorithm to optimize and gradient update (Hirose et al., 1991). Krizhevsky et al. (Krizhevsky et al., 2012) proposed convolutional neural network model that won the 2012 Imagenet model and this research eventually give the rise of CNN usage in various image classification applications.

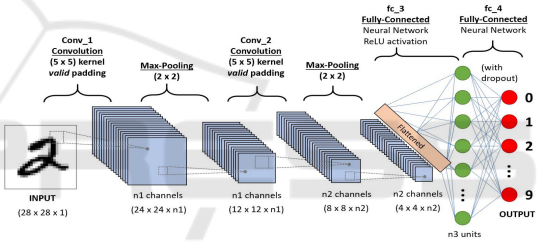

Figure 3: A Common CNN architecture, build with Keras (Ashiquzzaman and Tushar, 2017).

In the Figure 3, The common architecture of a classifying CNN is shown. Based on the given labels and the input, the whole model will learn to representation of images and map the features to the given label.

4.2 Activation Functions

In neural network, it is vitally important to use proper activation function. As the proposed model is mainly used convolutional layers for classifying into proper MOS category, ReLu (Rectified Linear Unit) has been selected for it's linear characteristics. As shown in the Equation 1, the Relu function gives the proper linear output for the given input and the function is easier to calculate in computational perspectives.

$$Relu(p) = \begin{cases} 0 & for\ p < 0 \\ 1 & for\ p \geq 0 \end{cases} \quad (1)$$

4.3 Regularization or Batch Normalization

Batch normalization reembersed gradient flow through the neural net which allows higher learning rate. moreover, it reduces strong dependence on initialization, because initial values even out due to normalization within batch. Essentially, batch normalization is a form of regularization where representation of an input at some layer is tied with other inputs in the same batch. Therefore, output of a node now turns into the function of input and other inputs in same batch. This in turn helps in generalization in representation space on that layer (Ioffe and Szegedy, 2015).

Algorithm 1: Batch Normalization by Ioffe and Szegedy (Ioffe and Szegedy, 2015).

procedure BATCH NORMALIZATION(x)
  Input: $x \rightarrow \mathcal{B} = \{x_1, \cdots, x_m\}$
  Output: $y_i = BN_{\gamma,\beta}(x_i)$
  $\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i)$ mini-batch mean
  ${\sigma_{\mathcal{B}}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2$ mini-batch variance
  $\hat{x_i} \leftarrow \frac{(x_i - \mu_{\mathcal{B}})}{\sqrt{{\sigma_{\mathcal{B}}}^2 - \varepsilon}}$ normalize
  $y_i \leftarrow \gamma x_i + \beta \equiv BN_{\gamma,\beta}(x_i)$ scale and shift
end procedure

Here the algorithm 1 by Ioffe and Szegedy is explained above.

4.4 Proposed Network Configuration

The proposed Neural Network model that used for the assessment tool is proposed in the following way. As the whole CNN has to process the images and map the features according to the output MOS category. As discussed in the Section 3, the whole images were converted into distinctive image classes according to the MOS scores. In The Table 1, The whole model architecture was shown by layer wise configuration.

The main convolutional layers were added in various filter size based, but the main modification in this model were the batch normalization layer distribution in each models to ensure proper unbiased learning. Later the model were pooled and the densely connected layers mapped and merged the whole model to classify. The main main input layer takes the imgages as the input. This layer follows by another same 64 filters convolution layer with 3×3 kernel and followed by a simple 2×2 Maxpolling to reduce the calculation overhead. Layers above the previous described had the same configuration, but instead of 64 filters in each convolutional layer, it consisted of 32 filters. The output layers has 4 dense output. the other dense layers were configured with $[150, 30]$ neurons in each hidden layers. The whole model was designed to classify the images based on the MOS category it originally belongs.

Table 1: Architecture of Proposed Deep Neural Network.

Layer	Number of Neurons	Activation Function	Kernel Size
Convolutional	64	ReLu	3x3
Convolutional	64	ReLu	3x3
Max Pooling	N/A	N/A	2x2
Batch Normalization	N/A	N/A	N/A
Convolutional	32	ReLu	3x3
Convolutional	32	ReLu	3x3
Max Pooling	N/A	N/A	2x2
Batch Normalization	N/A	N/A	N/A
Convolutional	32	ReLu	3x3
Convolutional	32	ReLu	3x3
Max Pooling	N/A	N/A	2x2
Batch Normalization	N/A	N/A	N/A
Convolutional	32	ReLu	3x3
Convolutional	32	ReLu	3x3
Max Pooling	N/A	N/A	2x2
Fully Connected	150	ReLu	N/A
Fully Connected	30	ReLu	N/A
Output	4	Softmax	N/A

The CNN model used in the based on the proposed model is simplified and shown in the Fig. 4. The experimental model was build based on this diagram.

5 EXPERIMENTS AND RESULT ANALYSIS

We employ FNCP (Future Network Computing Platform) to build a video delivery streaming system. The FNCP is built based on the NFV (Network Function Virtualization) proposed by ETSI (European Telecommunications Standards Institute). Thanks to its virtualized functions, we can quickly place a database server or a web server within a short time (in minutes). Besides, we can also reorganize or reconstruct network functions if we want to add or remove a component aiming to improve streaming quality. This task was intensively in the past because all of the tasks were done/processed with physical machines (computers).

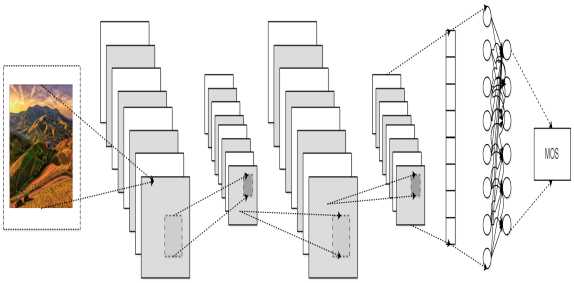

Figure 4: Proposed CNN model.

The QoE assessment model CNN was deployed in the python based program module in the machine. Experimental model was implemented in Python using Tensorflow and Keras libraries. The model is trained for 100 epoch with different initalization. The batch size is 128 for both training and testing. Categorical crossentropy is used as the loss function in this model. Adadelta optimizer (Zeiler, 2012) is used to optimize the learning process. Among the 10300 images, 9500 are used for training and 700 is used in validation. We have used a computer with CPU intel i7-9700U CPU @ 4,30GHz and @ 32GB RAM. Nvidia Geforce GTX 1050 ti dedicated graphics is used for faster computation, i.e CUDA support for accelerated training is adopted. The loss reduction of the CNN can be seen in the Fig.5 below.

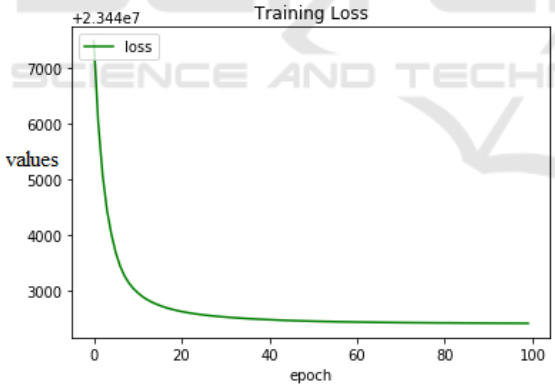

Figure 5: CNN loss visualization during training.

The resulted assessment model successfully labeled the images based on it's learning and the whole model achieved a 78% accuracy over whole augmented learning. The previous all methods described models with more accuracy, although all other methods had pre-computed layers or some modern technique is included. This experiment process has the raw RGB hi-def image as input and the whole model were trained to classify only by assessing the pixel values. The proposed model is thus the novel representation of the QoE assessment tools and had the state of the art accuracy.

6 CONCLUSIONS

4K or ultra-high-definition (UHD) will be the standard for video streaming in the next decade. In this research, we carry a study on deeplearning CNN QoE assessment tool which is a crucial adaptive algorithm. More specifically, we employ deep learning CNN algorithm in the form of quality of experience (QoE). In fact, QoE is an important factor to evaluate the efficiency of streaming transmission models. In this research article, we mainly focus on QoE performance analysis of streaming models to improve and ensure high quality (4K and UHD) streaming service in NFV. The developed CNN was state of the art QoE assessment tool with state of the art accuracy. The proposed model was crucial in the NR-VQA assessment model building up. In the future work, the main focus of this research will be applying some Autoencoder based data reduction model to reduce overhead and then apply it to the main CNN model for faster assessment.

ACKNOWLEDGEMENTS

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2018-2016-0-00314) supervised by the IITP(Institute for Information & communications Technology Promotion). Besides, this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (MEST)(Grant No. NRF-2017R1D1A1B03034429) and was also supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT). [2018-0-00691, Development of Autonomous Collaborative Swarm Intelligence Technologies for Disposable IoT Devices]. Finally, This work was supported by Electronics and Telecommunications Research Institute(ETRI) grant funded by ICT R&D program of MSIT/IITP[2200-2019-00008, Hyper-Connected Common Networking Service Research Infrastructure Testbed].

REFERENCES

Ashiquzzaman, A. and Tushar, A. K. (2017). Handwritten arabic numeral recognition using deep learning neural networks. In *Imaging, Vision & Pattern Recognition (icIVPR), 2017 IEEE International Conference on*, pages 1–4. IEEE.

Hirose, Y., Yamashita, K., and Hijiya, S. (1991). Back-propagation algorithm which varies the number of hidden units. *Neural Networks*, 4(1):61–66.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456.

Jung, R., Kornhuber, H., and Da Fonseca, J. S. (1963). Multisensory convergence on cortical neurons neuronal effects of visual, acoustic and vestibular stimuli in the superior convolutions of the cat's cortex. *Progress in brain research*, 1:207–240.

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

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

LeCun, Y. and Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995.

Lin, H., Hosu, V., and Saupe, D. (2018). Koniq-10k: Towards an ecologically valid and large-scale iqa database. *arXiv preprint arXiv:1803.08489*.

Nam, H., Kim, K.-H., and Schulzrinne, H. (2016). Qoe matters more than qos: Why people stop watching cat videos. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE.

Sheikh, H. R., Bovik, A. C., and De Veciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on image processing*, 14(12):2117–2128.

Van Ma, L., Ashiquzzaman, A., Kim, S. W., Lee, D., and Kim, J. (2018). Machine learning-based qoe performance analysis for dash streaming in nfv. In *Proceedings of KIIT Summer Conference*, pages 83–86.

Yoo, S.-J., Kwak, K.-S., and Kim, M. (2003). Predictive and measurement-based dynamic resource management and qos control for videos. *Computer Communications*, 26(14):1651–1661.

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.