https://doi.org/10.2352/ISSN.2470-1173.2020.11.HVEI-130

Predicting Single Observer's Votes from Objective Measures using Neural Networks

Lohic Fotio Tiotsop[1]; Tomas Mizdos[2]; Miroslav Uhrina[2]; Peter Pocta[2]; Marcus Barkowsky[3] and Enrico Masala[1]
[1]Computer Engineering Department, Politecnico di Torino (Italy) [2]University of Zilina (Slovakia), [3]Deggendorf Institute of Technology (Germany)

Abstract

The last decades witnessed an increasing number of works aiming at proposing objective measures for media quality assessment, i.e. determining an estimation of the mean opinion score (MOS) of human observers. In this contribution, we investigate a possibility of modeling and predicting single observer's opinion scores rather than the MOS. More precisely, we attempt to approximate the choice of one single observer by designing a neural network (NN) that is expected to mimic that observer behavior in terms of visual quality perception. Once such NNs (one for each observer) are trained they can be looked at as "virtual observers" as they take as an input information about a sequence and they output the score that the related observer would have given after watching that sequence. This new approach allows to automatically get different opinions regarding the perceived visual quality of a sequence whose quality is under investigation and thus estimate not only the MOS but also a number of other statistical indexes such as, for instance, the standard deviation of the opinions. Large numerical experiments are performed to provide further insight into a suitability of the approach.

Introduction

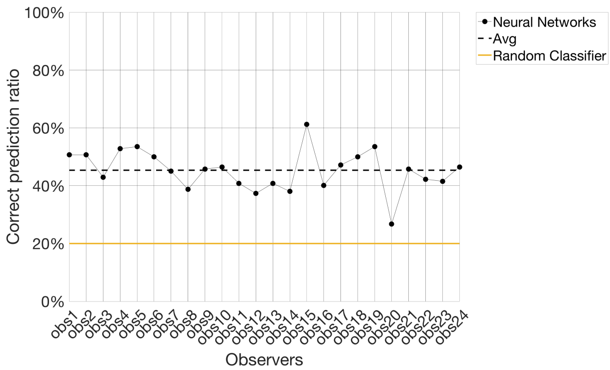
Machine learning models and algorithms have demonstrated state of the art performance in a large number of research fields. Traditionally, the use of these models in media quality assessment research is restricted to the prediction of the mean opinion score (MOS) [1–3] without considering other statistical parameters, e.g. a standard deviation and confidence interval. We aim at investigating new directions, i.e. approximating the opinions of an individual human observer through an artificial neural network (NN). By doing so, we expect to improve the design of subjective tests, hopefully making them more efficient. More precisely, starting from the data collected during a subjective experiment, we design for each observer involved in the experiment a NN that is expected to learn the main features that contribute to determine his/her opinion after watching a content. Such a NN would usually require to mimic the human visual system. It would thus be a multi-stage Deep Neural Network with one part mimicking early vision and then further parts for the higher cognitive interpretations. However, in this paper, we simplify it to a very limited approach by using a small scale network in order to reduce the computational burden. The NN of each observer is trained by some objective measures computed on the content of the training set and the ground truth opinions actually expressed by the observers.

Once the NN for each observer receives as input values computed by the deployed objective measures on a certain sequence, it provides as an output five probabilities p_i with $i = 1, \ldots, 5$, where p_i represents the probability that the related observer would have voted i. The predicted opinion is then the one corresponding to the highest probability. It is worth noting here that the probability distribution determined by the values p_i with $i = 1, \ldots, 5$ allows us to capture an uncertainty that would be included in observers' opinions. Furthermore, it models an inability of human observer to repeat, in a deterministic way, his/her opinion regarding the perceived visual quality of sequence when he/she would be asked to give the rating again.

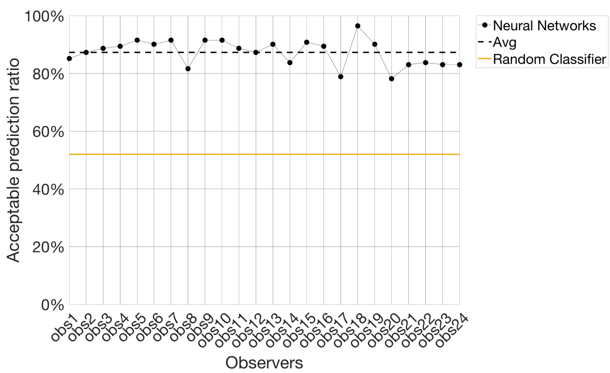
We rely on the VQEG-HD1, VQEG-HD3 and VQEG-HD5 experiments that involved in total 72 observers, i.e. 24 for each experiment. The sequences used in the VQEG-HD experiments have been chosen in order to consider a large amount of content, conditions, and visual perceived quality ranges. Hence, we believe that the dataset containing all the experiments mentioned above represents rather well the conditions that can be encountered in the majority of real-world applications. To be able to use more data for training and testing the NNs, we developed a method to form triplets of observers in order to obtain 24 "virtual" observers that we assumed to have attended all the 3 experiments. We then trained 24 NNs that are expected to mimic the visual quality perception of these 24 observers. We conducted large numerical experiments aiming first at investigating an ability of NNs to model single observer's opinions and then at assessing a suitability of the approach.

Motivation

Despite the fact that subjective experiments are expensive and time consuming, they are of paramount importance for a development and the validation of objective measures for media quality assessment. One of the issues when preparing a subjective experiment is to make sure that the subjectively perceived visual quality of the sequences to be prospectively deployed in the subjective test totally covers the chosen rating scale. In this work, as stated previously, we aim at mimicking the choice of a single observer by NNs, thus training "virtual observers". Such virtual observers can be used to simulate a subjective experiment on a set of sequences to have preliminary insights into a capability of such a set of the sequences to fully cover the MOS scale. Furthermore, each subject returns a discrete probabilistic distribution rather than a single opinion score (OS) as the NN modeling is deployed. So, it is possible to model, during the simulation, an inconsistency of human observers relying on such a distribution to perform the Monte Carlo simulation in order to get an accurate estimation of the MOS confidence intervals of each sequence and finally choose those that would best match the requirements.



(a) Percentage of correct predictions



(b) Percentage of acceptable predictions

Figure 1. *The models are trained on the VQEG-HD set 1 and 5 then tested on the VQEG-HD set 3 using only the VQMs as the features. For each observer, the related NN performs better than a random classifier*

The need of going beyond the MOS [4–7] in some situations is quite evident. For instance, additional statistical measures are derived in [4]. The usefulness of MOS distributions is in more details highlighted in [6, 7]. Moreover, a deployment of the MOS ranges instead of the single MOS values has been proposed in [5]. This need is also evident in practical use cases. For instance, streaming vendor companies need more than the MOS; in fact it is more important for them to make sure that a certain percentage of customers is perceiving a visual quality above a certain threshold. Hence being able to figure out an estimation of the distribution of the observers' opinions is of fundamental importance in this case. The NNs are trained to mimic a set of observers. The votes predicted by such NNs regarding the quality of a given sequence can serve as a sample to estimate the distribution of the opinions and hence to address the issues faced, for instance, by streaming vendors.

It is worth noting here that our approach, i.e. modeling single observers' opinions rather than the MOS, is much more general than a simple variance estimation as all the statistical indicators of subjective quality that have been considered in the literature can be estimated from the observers' votes. In other words, given a sequence, the designed NNs can be used to obtain different opinions as they model different observers. From these opinions one can estimate not only the MOS but also the standard deviation of the opinions (SOS) as well as the confidence intervals for the subjective quality or further indicators as desired.

Modeling single observer's opinions

As already mentioned above, we rely on the VQEG-HD experiments [8] to train and validate the NNs mimicking the behavior of observers in terms of the visual quality perception. For each processed video sequence (PVS) in the dataset, we computed a set of full-reference video quality measures (VQMs), more precisely PSNR, SSIM, MS-SSIM, VIF [9], and VMAF 0.6.2 [10], as well as 8 no-reference perceptual features, namely Blockiness, Blockloss, Blur, Noise, Contrast, Flickering, Spatial activity index (SI) and Temporal activity index (TI) [12]. We restrict our analysis to non-interlaced sequences, i.e. the VQEG-HD1, VQEG-HD3 and VQEG-HD5 subsets, as the most of the quality measures do not easily handle an interlaced video. It should also be noted here that the measures were not specifically designed for some distortion types included in the VQEG-HD dataset, for example transmission errors and temporal misalignments. For each of these three subsets, 24 observers participated in the experiment and expressed their opinions regarding the perceived visual quality of almost 160 sequences. Despite the fact that the three experiments took place in different laboratories, with different observers and different set of sequences, some sequences (3×24 sequences) were evaluated during all the experiments. In the rest of the work we refer to these sequences as the "common set".

In order to train a NN that models an observer in terms of visual quality perception, one should ideally consider only the sequences watched and evaluated by that observer. The NN should be then trained on such a data to approximate the process that the observer uses to assign his/her opinion. However, as previously pointed out, during the VQEG-HD experiments, each observer evaluated no more than 160 stimulus. Such a number of sequences is not large enough for effectively training and validating the NN that is supposed to model the choice of each observer. To have more data available for each observer, we approximate the observers' votes, concerning the sequences that have not been evaluated directly by that observer, by using those provided by another observer that voted similarly to the observer on the common set.

To be more precise, let's consider an observer $\hat{O}_1$ that participated in the VQEG-HD1 test, i.e. the experiment conducted on the VQEG-HD1 subset. For such a observer the votes for the sequences used during the VQEG-HD1 experiments are readily available. As this set of data is not enough to train and test the NN that would mimic such a observer, we estimate the votes that could have been given by the observer $\hat{O}_1$ to the sequences used in the VQEG-HD3 experiment by using those provided by an observer $\hat{O}_3$ that participated in the VQEG-HD3 experiment and voted very similarly to the observer $\hat{O}_1$ on the common set. The similarity is measured through the mutual residual mean square error between the votes given by the observers to the sequences included in the common set. Hence we find the observer $\hat{O}_3$ as follows:

$$\hat{O}_3 = \arg\min_{O_3} \sqrt{\frac{1}{|\mathscr{C}_{set}|}\left(\sum_{s \in \mathscr{C}_{set}} (V_s^{O_3} - V_s^{\hat{O}_1})^2\right)} \quad (1)$$

where O_3 indicates a generic observer of the VQEG-HD3 test, s a sequence, $V_s^{O_3}$ the vote of the observer O_3 for the sequence s and finally $\mathscr{C}_{set}$ represents the common set.

In the same manner, we related the observer $\hat{O}_1$ to an observer $\hat{O}_5$ of the VQEG-HD5 experiments. Hence starting from 72 observers, each having evaluated 160 sequences, we obtained 24 “virtual” observers by connecting each observers of the VQEG-HD1 test to one observer of the VQEG-HD3 as well as VQEG-HD5 tests that we assume having participated in all the 3 experiments and thus evaluated almost 500 stimuli.

We then trained the NNs, one for each virtual observer. Each NN was trained to predict the opinion score (OS) that the associated observer would have given after watching the sequence whose features were passed to the NN. We used five full reference metrics, i.e. PSNR, SSIM, MS-SSIM, VIF and VMAF as well as the 8 no-reference perceptual features listed above to train each NN. The NN modeling each observer learned a mapping from these features to the OSs of the related observer. At the end, 24 NNs were created/computed. Each one received as an input the aforementioned features computed on a given sequence and has predicted the OS that the observer modeled by such a NN would have been given after rating that stimulus.

Numerical experiments and results

The numerical experiments conducted in this section aim at: i) assessing the capability of NNs to effectively model a single observer in terms of visual quality perception; ii) Performing “virtual” subjective experiment using the 24 NNs that have been trained to mimic actual observers; iii) showing how our approach could be used to get insight into the uncertainty characterising the opinions of an observer.

For each of the 24 observers, we tested 3 different NN structures. More precisely, after setting the number of neurons of the input layer to the number of the features, we designed an output layer with 5 neurons, each one delivering the probability of each class of the 5-point rating scale. Finally, we tested 3 different configurations for the hidden layers, i.e. a single layer, two layers and finally three layers all having five neurons. We then chose, for each observer, the structure delivering the highest accuracy on the validation set. We observed an almost uniform distribution in terms of the number of layers, needed to model the virtual observers, providing the best accuracy.

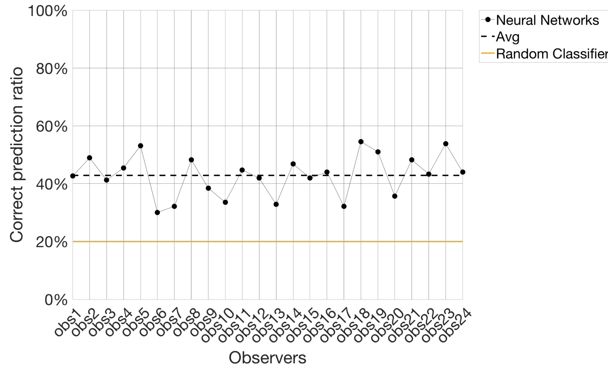
In order to assess the suitability of NNs to model the single observer’s opinions, we compared the performance of the 24 NNs to a random classifier (RC), i.e. a model that assign a random opinion to a given sequence. The aim of such an experiment is to demonstrate that the NN actually learned at least some aspects of the process that guides the choices of the observer, which it was trained to mimic. In fact, when the NN does not get any information about the visual quality perception of the related observer, it is expected to behave similarly to the RC and thus showing an accuracy not significantly different from the one of the RC.

We introduce the following 2 ratios:

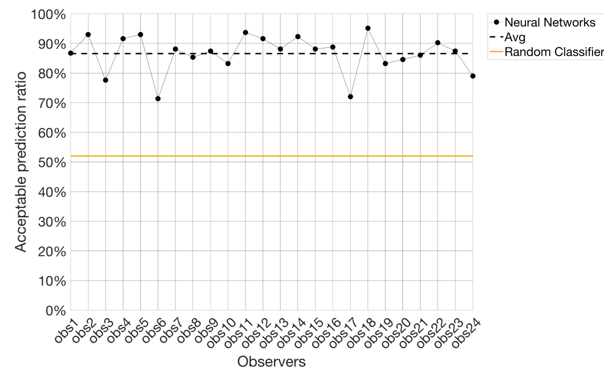
$$\textit{Correct prediction ratio} = \frac{\#(\textit{predicted OS=actual OS})}{\#(\textit{PVS in test set})}$$

$$\textit{Acceptable prediction ratio} = \frac{\#(|\textit{predicted OS-actual OS}| \leq 1)}{\#(\textit{PVS in test set})}$$

that are used to compare the accuracy of the 24 NNs to that of the RC. The correct prediction ratio or equivalently the accuracy of each NN achieved on a test set is a number of sequences for which



(a) Percentage of correct predictions



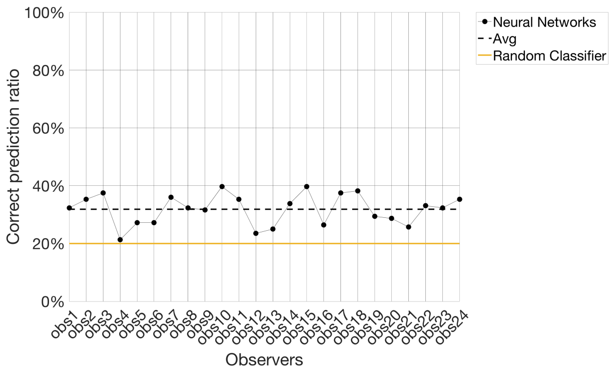
(b) Percentage of acceptable predictions

Figure 2. *The models are trained on the VQEG-HD set 3 and 5 then tested on the VQEG-HD set 1 using only the VQMs as the features. For each observer, the related NN performs better than a random classifier*

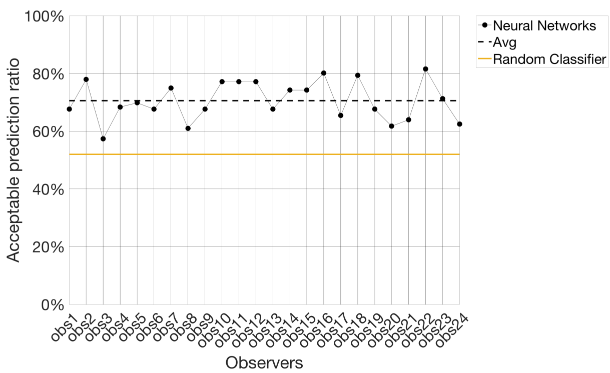
the opinion predicted by the NN is equal to the one given by the related observer divided by a total number of sequences in the test set. The acceptable prediction ratio is a number of sequences for which the NN prediction differs no more than 1 from the opinion of the related observer divided by the total number of sequences in the test set. For the RC, these ratios are respectively 20% and 52%, which are the expected values considering all the possible favorable cases (2/5·1/5+3/5·3/5+2/5·1/5).

We first trained the 24 NNs by using only the VQMs as the features. To assess the superiority of the 24 NNs against the RC or equivalently to show that each NN actually learned some information regarding the perception of quality of the modeled observer, we cross tested the NNs on the 3 VQEG-HD experiments. To be more precise, we trained the NNs on the data coming from 2 experiments and tested them on the remaining one. The results are shown in Figure 1, Figure 2 and Figure 3 respectively. In all the cases, the NN of each observer provides a higher accuracy than the RC and the average accuracy as well as the average ratio of acceptable predictions of the NNs overcame the one of the RC. This clearly implies that the NNs of each observer did learn something from the data and thus has modeled some aspects of the way the related observer perceives and judges quality.

We then trained the NNs taking into account the no-reference perceptual features listed above as well as their standard deviations over the frames of a sequence in addition to the VQMs. The



(a) Percentage of correct predictions



(b) Percentage of acceptable predictions

Figure 3. *The models are trained on the VQEG-HD set 1 and 3 then tested on the VQEG-HD set 5 using only the VQMs as the features. For each observer, the related NN performs better than a random classifier*

NNs were trained on the data of the VQEG-HD experiments 1 and 5 and tested on the data of the experiment 3. The Figure 4 reports the performance of the NNs, see Figure 4a and Figure 4c for the the training set, and Figure 4b and Figure 4d for the test set. As expected, more features led to the higher accuracy on average for the training set as more complex structures of NNs were considered. However when using such NNs on the test set, we observed the lower accuracy compared to the NNs trained only with the VQMs. This last observation is important as it implies that the accuracy of the NNs could be improved should we have enough data to consider more complex NNs structures. However, Figure 4 shows that, when considering the test set rather then the training one, the average performance of the NNs is not that far from the one obtained for the training set even though only the 5 VQMs were used as the input.

We now present the results that we obtained when using the 24 NNs as "virtual observers" in a "virtual" subjective experiment. Also for this experiment, we consider, as the training set, the VQEG-HD experiment 1 and 5. More precisely, we used the 24 NNs on each sequence of the VQEG-HD experiment 3. Hence, 24 predicted opinions were available for each sequence. We then computed, for each sequence, the average and standard deviation of the opinions provided by the 24 NNs that we named respectively "virtual MOS" and "virtual SOS". Figure 5 and Figure 6 show respectively the comparison of the MOS values and the "virtual MOS" values and the SOS values and the "virtual SOS" values. The results are quite promising as very high correlations were obtained for the MOS and the "virtual MOS" values. Furthermore, we obtained correlation coefficients significantly different from 0 between the SOS and the "virtual SOS" with 95% of confidence; this is encouraging as we are not aware of any other SOS estimation method that led to predictions significantly correlated to the SOS. Anyhow, it is worth noting here that a practical usage is however still limited as the Pearson correlation only reaches 0.5.

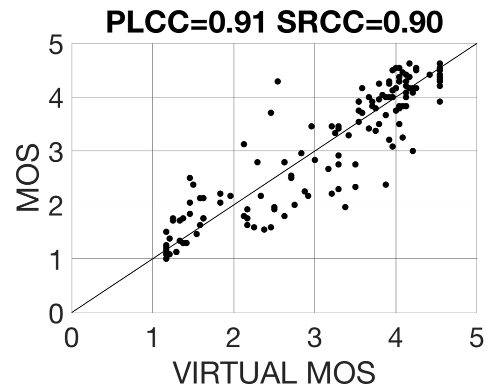


Figure 5. *The NNs (trained on the VQEG-HD set 1 and 5) are considered as "Virtual Observer" and used to run a "Virtual subjective experiments" on the VQEG-HD set 3. The obtained MOS values were compared to the actual ones*

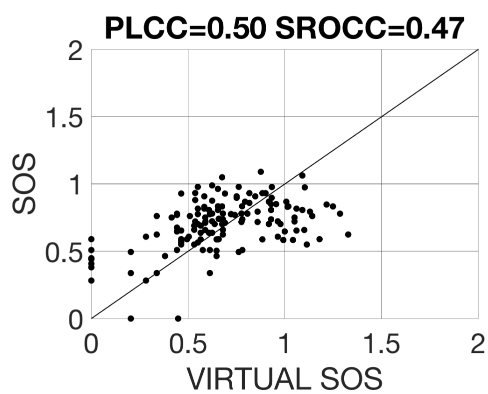
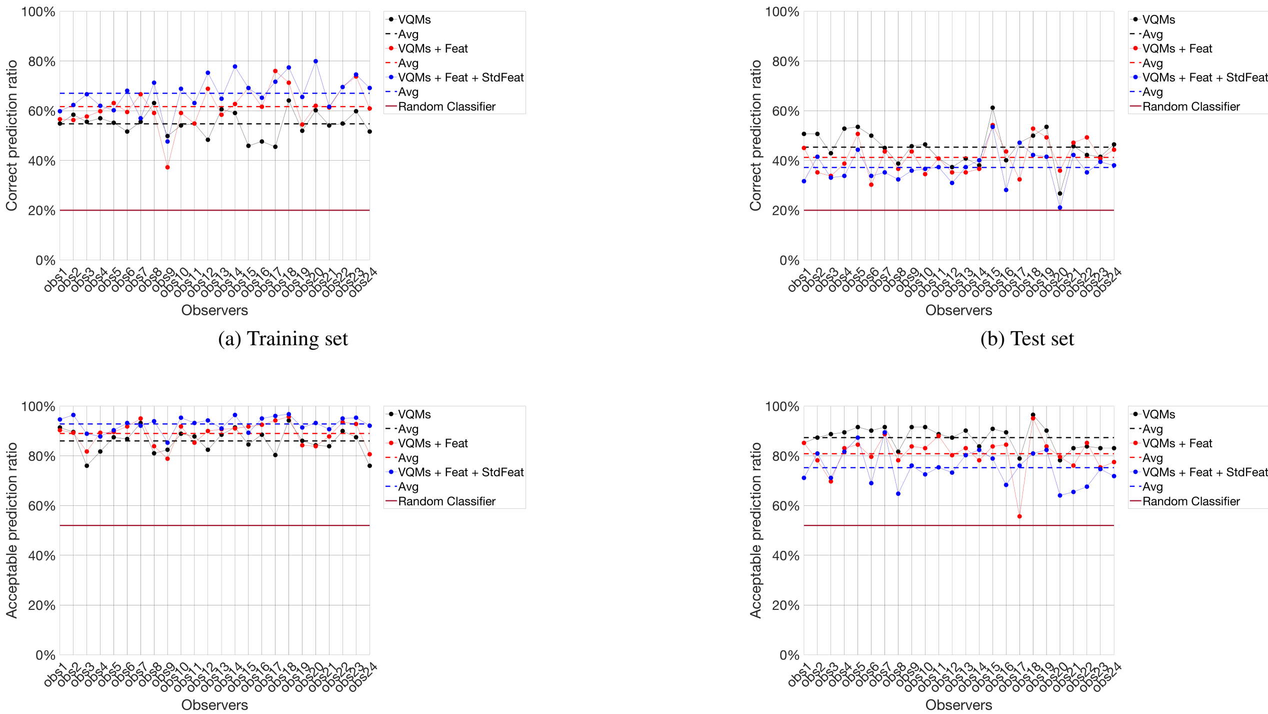


Figure 6. *The NNs are considered as "Virtual Observer" and used to run a "Virtual subjective experiments". The obtained SOS values are compared to the actual ones. The correlation coefficient is significantly different from 0* ($p_{value} = 2.1 \times 10^{-10}$)

We remind that the NN mimicking each observer outputs a discrete probabilistic distribution, p_i with $i = 1, 2, \ldots, 5$, on the 5-point absolute rating category scale when assessing the quality of a sequence. While the opinion of the observer is chosen as the mode of such a distribution, i.e. the opinion score with the highest probability, the variance of that distribution can be used to measure the uncertainty of the observer regarding the perceived visual quality of the sequence whose quality is under investigation. By the uncertainty of the perceived visual quality we mean, for instance, an inability of the observer to provide exactly his/her



(a) Training set (b) Test set

(c) Training set (d) Test set

Figure 4. *Contribution of no-reference perceptual features on the training (left) and validation (right) set, when predicting the OS.*

previous opinion again when asked to evaluate the same sequence again after a certain amount of time. Hence, after assessing the quality of the sequence s, using the NN mimicking the observer o, we define the following variance

$$\sigma_o^s = \sum_{i=1}^{5} i^2 \cdot p_i - \left(\sum_{i=1}^{5} i \cdot p_i \right)^2$$

as the uncertainty or the inconsistency of the observer o concerning the visual quality of s or equivalently a measure of the inability of o to provide the same opinion upon many ratings of the sequence s.

We also designed a numerical experiment aiming at determining which perceptual features among those considered in this work contribute the most to the uncertainty or inconsistency of an observer when it comes to the perceived visual quality of a sequence. We relied on the neighborhood component analysis (NCA) feature selection approach. The NCA allows to figure out how important would be a given feature when used to predict a given target variable. Further insights into the NCA can be found in [11]. We computed the importance of each of the 8 no-reference perceptual features mentioned above when predicting the uncertainty, i.e. the values σ_o^s for each observer. The result of the analysis is reported in Figure 7. For each observer, a height of the rectangle associated with each feature represents how important is such a feature in determining the uncertainty of the observer when it comes to the perceived quality of a sequence after watching it. It can be seen from the corresponding figure that the noise feature seems to be less important than the presence of blocking artefacts in the content when determining the capability of an observer to repeat his/her previous rating on a given stimulus. Furthermore, the observer #24 seems to exhibit a particular behavior. In fact, while the blur feature influences significantly the inconsistency of the other observers, his/her judgment is not at all influenced by the presence of the artefacts due to blur.

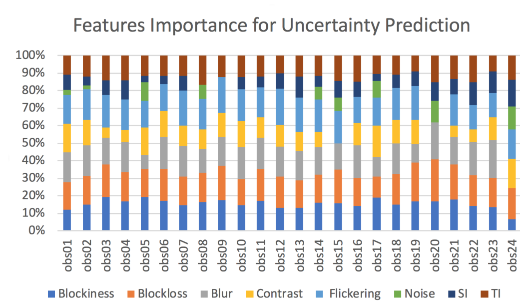
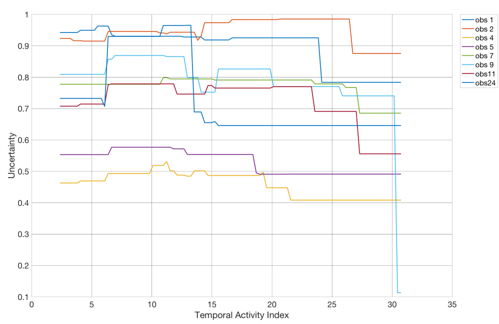
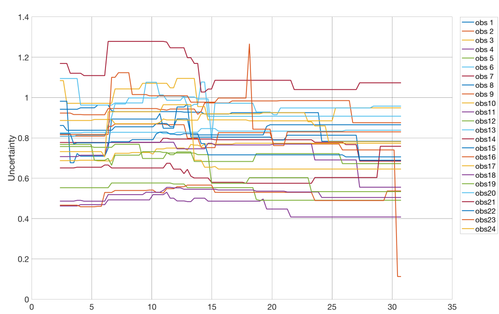


Figure 7. *Importance of (no-reference) features for the uncertainty prediction using the neighborhood component analysis (NCA). It seems that the noise feature is less influential than blockiness when it comes the uncertainty.*

Finally, we assessed the ability of the "virtual observers" to reproduce the content masking effect, i.e. the reduction or emphasis of distortion visibility due to the quantity of motion in the video sequence whose quality is under investigation. To that aim, we fitted the 8 no-reference perceptual features to the uncertainty values for each observer using a regression tree and then extracted the partial dependency of the uncertainty on the temporal activity index (TI) of the video sequence. Such a dependency represents the averaged predicted response of the regression tree as a function of the TI input. The results are shown in Figure 8. We show the curves, i.e. a relation between the TI of a video sequence and the uncertainty or the inconsistency of an observer when it comes to the perceived quality of the corresponding sequence, for all the observers as well as for some specific observers to better visualize the main outcome of the analysis. A closer look at the curves indicates that a lower value of inconsistency is observed for each observer when it comes to the sequences with a high value of TI. We attribute this behavior to the fact that temporal activity could mask part of the distortion that could contribute to a rise of the inconsistency of an observer. This may be seen as a first indication that the designed virtual observers can mimic this aspect of the visual system of real observers. It is worth noting here that a rather large gap is sometimes present between the curves of different observers. For instance, the model for the observer 2 seems to be more inconsistent than that of the observer 4, which may allow for some conclusion on the observer's behavior, but further investigation is required.



(a) Some observers



(b) All observers

Figure 8. *Lower uncertainty values are observed for sequences with a high temporal activity. Curves are obtained by fitting the perceptual features to the uncertainty by the regression tree models*

Conclusion

In this work, we investigated a different direction for media quality assessment using NNs. Rather than using them to build a model for MOS prediction as largely done in the literature, we suggest to employ them to model a single observer in terms of visual quality perception. So, we trained NNs that can mimic the behavior of real observers, thus taking as an input the information about a sequence and predicting as an output the rating that the observer would have provided after watching that stimulus. The preliminary results show that NNs can effectively model single observer's behavior. In fact, our trained NNs, used on a set of contents that has not been included in the training process, have been able to reproduce, with a significant accuracy, the MOS of the sequences as rated by the corresponding observers. Furthermore, our approach allowed us to evaluate also the uncertainty of an observer when it comes to his/her rating regarding the quality of a given stimulus as well as an impact of the perceptual features on it. The results reveal some diversities among the observers, hence supporting the very common idea that a visual quality perception of multimedia content is also influenced by variables that are not directly observable or measurable.

Acknowledgment

This work has been supported in part by the Politecnico di Torino Interdepartmental Center for Service Robotics (PIC4SeR) https://pic4ser.polito.it.

References

[1] D. Varga, T. Szirányi, No-reference video quality assessment via pretrained CNN and LSTM networks. Signal, Image and Video Processing, 13, pp. 1569-1576 (2019)

[2] C. G. Bampis, Z. Li and A. C. Bovik, Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment, IEEE Transactions on Circuits and Systems for Video Technology, 29, 8, pp. 2256-2270 (2019).

[3] J. You and J. Korhonen, Deep Neural Networks for No-Reference Video Quality Assessment, Proc. IEEE Intl. Conference on Image Processing (ICIP), pp. 2349-2353 (2019).

[4] T. Hoßfeld, P.E. Heegaard, M. Varela, S. Möller. QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS, Quality and User Experience, 1, 2 (2016).

[5] L. F. Tiotsop, E. Masala, A. Aldahdooh, G. V. Wallendael and M. Barkowsky, Computing Quality-of-Experience Ranges for Video Quality Estimation, Proc. Intl. Conference on Quality of Multimedia Experience (QoMEX) (2019).

[6] M. Seufert, Fundamental Advantages of Considering Quality of Experience Distributions over Mean Opinion Scores, Proc. Intl. Conference on Quality of Multimedia Experience (QoMEX) (2019).

[7] L. Janowski, Z. Papir, Modeling subjective tests of quality of experience with a Generalized Linear Model, Proc. Intl. Conference on Quality of Multimedia Experience (QoMEX) (2009).

[8] VQEG, Report on the validation of video quality models for high definition video content (v. 2.0), http://bit.ly/2Z7GWDI (2010).

[9] P. Hanhart and R. Hahling, Video Quality Measurement Tool (VQMT), http://mmspg.epfl.ch/vqmt (2013).

[10] Netflix, VMAF - Video Multi-Method Assessment Fusion, https://github.com/Netflix/vmaf (2019).

[11] W. Yang, K. Wang, W. Zuo, Neighbourhood Component Feature Selection for High-Dimensional Data, Journal of Computers, 7, 1, pp. 161-168 (2012).

[12] M. Leszczuk, M. Hanusiak, M.C.Q. Farias, E. Wyckens, G. Heston, Recent developments in visual quality monitoring by key performance indicators, Multimedia Tools and Applications, 75, 10745–10767 (2016).

Author Biography

Lohic Fotio Tiotsop received his M.Sc. degree in Mathematical Engineering from Politecnico di Torino, Italy. Since 2018 he is a Ph.D. stu-

dent in Control and Computer Engineering at Politecnico di Torino. His primary research interests are advanced statistical methods and machine learning algorithms applied to multimedia problems.

Tomas Mizdos received his B.Sc. degree in Multimedia Technologies from the University of Zilina in 2015. He is currently a Ph.D. student in Telecommunications at the University of Zilina. His main areas of interest are functionality and quality of multimedia services, quality of experience, video coding and digital signal processing.

Miroslav Uhrina received his Ph.D. degree in Telecommunications at University of Zilina in 2012, where he is now assistant professor. His research interests include quality of experience, audio and video compression, TV broadcasting and IP networks.

Peter Pocta received his Ph.D. degree in 2007 from the University of Zilina, where he currently serves as an Associate Professor at the Department of Multimedia and Information-Communication Technology. He is involved in International Standardization through the ETSI TC STQ as well as ITU-T SG12. His research interests include speech, audio, video and audiovisual quality assessment, speech intelligibility, multimedia communication and QoE management.

Marcus Barkowsky received his Dr.-Ing. degree from the University of Erlangen-Nuremberg in 2009. He joined the University of Nantes in 2010, then in 2018 he obtained the professorship on interactive systems and internet of things at the Deggendorf Institute of Technology, University of Applied Sciences. His activities range from designing 3-D interaction and measuring visual discomfort using psychometric measurements to computationally modeling spatial and temporal effects of the human perception.

Enrico Masala received his Ph.D. degree in Computer Engineering in 2004 at the Politecnico di Torino, where he is currently associate professor. His main research interests include multimedia quality optimization of communications over packet networks, with special attention to particular scenarios such as remote control applications, 3D video, cloud for multimedia. He is also involved in the management of the Politecnico di Torino Interdepartmental Center for Service Robotics (PIC4SeR).