Deep Learning for Quality Assessment in Live Video Streaming

Maria Torres Vega, *Student Member, IEEE,* Decebal Constantin Mocanu, *Student Member, IEEE,* Jeroen Famaey, *Member, IEEE* Stavros Stavrou, *Senior Member, IEEE* and Antonio Liotta, *Senior Member, IEEE*

Abstract—Video content providers put stringent requirements on the quality assessment methods realized on their services. They need to be accurate, real-time, adaptable to new content, and scalable as the video set grows. In this letter, we introduce a novel automated and computationally efficient video assessment method. It enables accurate real-time (online) analysis of delivered quality in an adaptable and scalable manner. Offline deep unsupervised learning processes are employed at the server side and inexpensive no-reference measurements at the client side. This provides both real-time assessment, as well as performance comparable to the full reference counter-part, while maintaining its no-reference characteristics. We tested our approach on the LIMP Video Quality Database (an extensive packet loss impaired video-set) obtaining a correlation between 78% and 91% to the FR benchmark (the Video Quality Metric, VQM). Due to its unsupervised learning essence, our method is flexible, dynamically adaptable to new content and scalable with the number of videos.

Index Terms—Deep learning, unsupervised learning, video quality assessment, multimedia video services.

I. INTRODUCTION

ASSESSING video quality has been traditionally performed by means of subjective or complex objective Quality of Experience (QoE) metrics [1], due to their demonstrated correlation to human visual perception [2], [3]. However, these come with stringent computational complexity and time requirements. They are therefore inapplicable in situations where real-time analysis is needed. Examples of such situations include adaptive streaming systems and real-time quality of experience management [4], [5], [6]. Furthermore, as new video streaming systems, compression standards, and content types appear, the scalability of the assessment methods becomes crucial.

Reduced-Reference (RR) and No-Reference (NR) metrics are best suited for real-time evaluation. They assess quality purely by means of specific features extracted from the received video signals in combination with the network conditions [7]. This is a very difficult task. Consequently, most of these metrics focus on the specific behavior of particular distortions, such as the level of blur or noise within the frames,

M. Torres Vega, D.C Mocanu and A. Liotta are with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands. J. Famaey is with the University of Antwerp-iMinds, Antwerp, Belgium. S.Stavrou is with the Faculty of pure and Applied Sciences, Open University of Cyprus, Nicosia, Cyprus

to make their assessment. Examples of these are the frame freezing approaches of Huynh et al. [8] and Mok et al. [9] or the generalized local binary pattern approach for image quality assessment of Min Zhang et al. [10]. However, this type of metrics fails to provide an accurate assessment when more than one artifact is present. For this reason, nowadays there is a trend of metrics that try to combine the effect of more than one artifact to provide more accurate measurements. Image or video statistics modeling has been considered for developing quality metrics [11]. Zeng et al. proposed the use of temporal motion smoothness of a video sequence to examine the temporal variations of the local phase structures in the complex wavelet transform domain [12]. Other approaches have focused on trying to model the distortion based on the encoding of the video sequences. Examples are the MPEG-2 spatial and temporal features extraction of Wolf et al. [13] or the DCT measurement of Yan et al. [14]. These approaches, while providing good results on compression derived distortion, are unfit for real-time network impaired videos, due to their high computational and time complexity.

By means of prediction, which can be trained offline at the server-side, the accuracy of the assessment can be improved without increasing its complexity on the client side. Promising examples of cognitive approaches are adaboost for assessing artifact levels in videos [15], the bitstream based artificial neural network [16], the artificial neural network for jerkiness evaluation [17], and the regression framework for estimating the objective quality index [18].

A key limitation of current studies is that they have focused on compressed or synthetically-distorted videos. Furthermore, the majority of learning-based approaches have used supervised learning techniques, which train a prediction model on labeled samples, based on the ground truth quality, subjective or full reference index. This process obviously slows down the assessment procedure. Moreover, it scales poorly as the introduction of new video types in the system and distortion conditions in the network requires manual full subjective reference sample labeling. The aim of our work is to introduce a method that can work in general cases, particularly in real-time streaming, under realistic network distortions and in a scalable manner. This is fundamental in real-time transmission systems [19] and allows the solution to deal with the broad amount of video types and streaming conditions, which are typically unknown at design time. To achieve this, we have taken the unsupervised deep learning (UDL) path.

Among all the available UDL techniques, the Restricted Boltzmann Machines (RBM) [20] have been successfully

applied to still images in previous work [21], due to their outstanding performance as density estimators. In this work, we combine this type of learning with light-weight NR metrics to obtain a video quality assessment comparable in accuracy to the FR state-of-the-art. We tested our method on the LIMP Video Quality Database [22], [23], an extensive packet-loss impaired video-set, benchmarking our results against the Video Quality Metric (VQM) [24], due to its demonstrated correlation to the human visual system [3]. Correlations between 78 and 91% were obtained, depending on the video and network conditions. Furthermore, our method requires only the original videos to accurately assess all the compressed and network impaired subsets. In this way, our work presents an accurate, real-time, and adaptable video quality method suited for sizable video sets.

In the remainder of this letter, we provide a short introduction to unsupervised learning and Restricted Boltzmann Machines (Section II). Further on, Section III presents the video quality method that is then evaluated in Section IV. We conclude our paper and give directions for future work in Section V.

II. UNSUPERVISED LEARNING, DEEP LEARNING AND RESTRICTED BOLTZMANN MACHINES

Adaptability, scalability and accuracy are crucial characteristics for a video service provider to select a quality assessment method. First, fast adaptability of the model when new videos enter the system is fundamental. If the model belongs to the supervised learning type (e.g., artificial neural networks, regression models), a newly released video sample needs to be manually labeled (its ground truth obtained) before inclusion. This action slows down the process and the adaptability requirement will not be achieved. For this reason, we turned to unsupervised learning (UL) methods. Second, to master the sheer scale of the problem, we selected deep learning (DL) techniques. Within the broad variety of DL techniques, the Restricted Bolztmann Machines (RBM) have demonstrated their usefulness as density estimators [21], [25]. We used RBMs for the design of our cognitive video quality assessment method.

UL is the machine learning task of inferring a function to describe the hidden structure of unlabeled data [26]. In recent years, several approaches have been used to enhance the prediction characteristics of this type of model. Among all, Deep Learning (DL) [27] is being actively used in problems where scalability is of prime importance. Deep learning attempts to model high-level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations. Among these models combining unsupervised and deep learning (Unsupervised Deep Learning, UDL), Restricted Boltzmann Machines [20] have shown outstanding performance as density estimators [21] [25] [28].

RBMs are two-layer generative stochastic artificial neural networks that can learn a probability distribution over its set of inputs by means of only inter-layer connections. During training, the inputs (visible layer) are translated into a higher
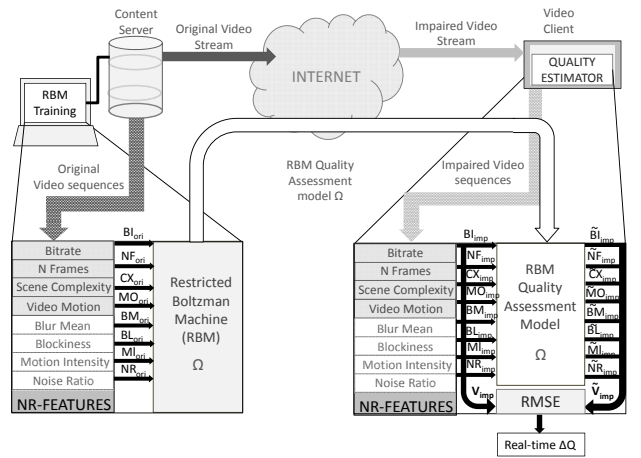

Fig. 1: Real-Time UDL-based Video Quality Assessment Method. Note that Server and Client processes are de-coupled; the server process (RBM training) runs in the background; and the model Ω is transferred to the client offline.

feature space (hidden layer) by means of inter-layer connections. This translation has the objective of minimizing the error between the inputs and the reconstructed inputs. After training, when the model encounters inputs not belonging to the learned distribution, the error between the inputs and their reconstructed versions increases [28]. In this work, we have built on this notion to estimate video quality degradation by means of the error distance between the inputs and their reconstructed versions. For it, the characteristics of the original video content act as visible neurons. Both, visible and the hidden neurons have an associated bias, which together with the inter-layer weights characterize the RBM model, Ω.

III. REAL-TIME COGNITIVE VIDEO QUALITY ASSESSMENT METHOD

In this section, we present our UDL-based method. Figure 1 shows the processes taking place both on the server (offline) and client (in real-time).

Like any other prediction-based method, ours requires a training phase which takes place at the server side in an offline manner. In it, the RBM model (Ω) is trained with the original video sequences available in the content delivery service. Each training sample is composed of eight NR features extracted from the corresponding original video sequence. These sequences are usually of a duration between 2 and 10 seconds, depending on the video provider. In this way, long available video content can be splitted in these smaller video sequences of a pre-established length, to be used for training the RBM. When the client session starts, the RBM model is transmitted to the client device, ready to be used when the streaming session takes place. If new video sequences are added to the content provider catalog, their features are extracted, the RBM model (Ω) is retrained (adding the new samples) and an update is sent to the client. An RBM model consists of the visible and hidden biased, as well as the interlayer weights. An update requires the transmission of $8 + M + 8 \times M$ real numbers, where M is the number of hidden neurons. The training of the

model proceeds independently from the real time assessment in the client; hence, our method falls within the NR category. In principle, our model could also be implemented as an RR metric, whereby the model parameters are passed to the client online. Yet, in this letter we have chosen the offline scenario due to its easier applicability in a real-life scenario.

On the other end of the transmission chain, while a streaming session is taking place, the client performs real-time extraction of the eight NR features on the stream. When the current video sequence is finished, the client averages the results. These eight values ($V_{imp} = [BI_{imp}, ..., NR_{imp}]$) serve as input to the RBM server model (Ω). The model outputs the estimated values corresponding to the trained model, i.e. the estimated values for the impaired version of the video ($\tilde{V}_{imp} = [\tilde{BI}_{imp}, ..., \tilde{NR}_{imp}]$). Finally the relative degradation (ΔQ) is calculated as the Root Mean Squared Error (RMSE) [29] of the impaired measured values (V_{imp}) and the RBM reconstructed values ($\tilde{V}_{imp}$). Through this process, our method measures the relative degradation between the original video and the one received after the transmission chain, without requiring the original video. Given the fact that the RMSE takes values in [0, 1), our method measurements range between 0 and 1, where zero indicates full quality and one means full degradation.

In order to formalize the RBM, it is crucial to choose appropriate learning rules to be used for fitting the input values into the model [30], [31], [32]. The most used approach is the Contrastive Divergence (CD) method proposed by Hinton [33], which performs an approximation of the maximum likelihood learning. The update number, learning rate, momentum, and weights decay together determine the learning rules [34]. Our solution makes use of this well-known learning method.

The NR features were selected relying on our previous study of the accuracy of simple NR methods in the presence of network impairments [23], focusing on the ones with demonstrated correlation to quality degradation. Including features that do not correlate with video quality degradation could negatively influence the UDL estimation process and therefore reduce accuracy. Based on those results, we selected six NR low complexity metrics, both in the pixel and the bitstream layers. Two additional video characteristics linked to quality degradation (namely the video bitrate and the number of frames) were also added to the features. This resulted in a total of eight NR features, as referenced above.

A video stream can be characterized by several parameters. First, the bitstream bitrate influences the quality in a direct and substantial manner. Higher bitrates result in higher quality indices [35]. Second, the received number of frames gives an idea of the duration of the video. Finally, parameters regarding the video scene composition have been demonstrated to affect quality to a large extent [36]. From these parameters, the scene complexity and the video motion have proven to give a high correlation with video quality [23]. Both these features can be empirically obtained from the encoding [36]. These four features can be obtained from the stream as it finishes and are directly computed.

On the pixel level, we demonstrated [23] that in video streaming, degradations in terms of the level of blockiness [37], [38], the noise ratio and the average blur [39] are well correlated with the overall quality index. In addition to this set, we added a feature concerning the temporal characteristics of the video on the pixel level, the motion intensity, which measures the movement of video objects between frames by means of the compared level of intensity [40]. The reasoning behind selecting these features and not others comes from the need to pursue low computation and ability to be performed in real-time even on light-weight devices such as smartphones and tablets, as we demonstrated in previous work [41], [23].

IV. EXPERIMENTAL EVALUATION

The LIMP Video Quality Database [22], [23] was used to evaluate our proposed model. For it, the RBM model in the server is trained with the original videos sequences of this set (9 different video sequences) provided by Seshdrinathan et al. [42] and encoded using MPEG-4 part 10/H.264. Subsequently, we tested the client with each of the network and compression impaired videos of the set.

The LIMP dataset [22] consists of 9 high quality videos (i.e., bs1, mc1, pa1, pr1, rb1, rh1, sf1, sh1, tr1) from the Live Quality Video Database [42]. Each video has a frame rate of 25fps, a duration of 10 seconds and a resolution of 768x432. This resolution was selected in the original database due to the fact that it ensures that the aspect ratio of the original raw videos (taken with high definition cameras) was maintained, thus minimizing visual distortion while adapting videos to resolution constrained environments, such as smartphone screens. These original nine videos were encoded using MPEG-4 part 10/H.264 at 8 bit rate levels (i.e., 64, 640, 768, 1024, 2048, 3042, 4096, and 5120 kpbs) obtaining 72 unimpaired videos. Each of these (9 videos at 8 bitrates each) was then streamed in a controlled network environment (using the PacketStorm Hurricane II network emulator [43]) and subjected the videos to 12 levels of randomized packet loss (i.e., 0%, 0.5%, 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, and 10%). This makes a total of 864 different videos on which to assess the accuracy of our UDL-based quality method. This video set allows for an extensive analysis of the effects that packet loss has on different video types. We focused on packet loss effects, due to its being the most impairing network condition [4], [19]. Our approach calculates the average quality over the entire video fragment. As such, it takes into account the loss of every type of frame, and statistically averages loss probability over time.

The RBM is structured with 8 visible neurons (one per feature) and 50 hidden neurons. These settings make a total of 458 real numbers, the RBM free parameters (8 visible and 50 hidden biases plus 400 interlayer weights). This amounts to roughly 1.5KBytes to send between server and client when an update is required. Based on insights from previous work [33], the learning rate was set to 0.01, the number of CD steps to 1, the weight decay to 0.0002, the momentum to 0.9, and we trained the models for 100 epochs. The RBM was then trained on the 9 original high quality videos, while the other 855 variations were used for testing.

We evaluated the performance of our method as a relative degradation metric by means of three correlation measure-

TABLE I: Results of the two experiments where values indicate PCC, P-value an RMSE averaged for each video type across all compression levels. Columns 2 to 4 belong to the Compression experiment (Section IV-A). Columns 5 to 7 refer to the Network loss experiment (Section IV-B). Cell colors give qualitative correlation levels: green (best), yellow (medium), and red (worst)

V.T.	Compression			Network Loss		
	PCC	P-VAL	RMSE	PCC	P-VAL	RMSE
bs1	0.99	<0.001	0.08	0.69	0.07	0.1
mc1	0.98	<0.001	0.13	0.79	0.02	0.20
pa1	0.84	0.009	0.17	0.82	0.002	0.20
pr1	0.93	<0.001	0.23	0.68	0.07	0.25
rb1	0.93	<0.001	0.20	0.77	0.2	0.20
rh1	0.64	0.08	0.17	0.85	0.001	0.20
sf1	0.92	<0.001	0.14	0.85	<0.001	0.12
sh1	0.99	<0.001	0.22	0.78	0.05	0.25
tr1	0.99	<0.001	0.19	0.78	0.02	0.22
All	0.91	≈0.01	0.17	0.78	≈0.03	0.19
	±0.11	±0.03	±0.05	±0.06	±0.03	±0.05

ments, using as benchmark the well-known FR degradation assessment model VQM [24], due to its demonstrated correlation to human perception [3]. The difference between our metric and the benchmark is measured by the Root Mean Square Error (RMSE). Linear correlations are assessed by the Pearson Correlation Coefficient (PCC). Finally, the P-value shows if statistically our UDL-approach comes from the same distribution as the benchmark FR quality. These three measurements are presented by means of per-video-averages (Table I) and per bitrate and video type colormaps (Figure 2).
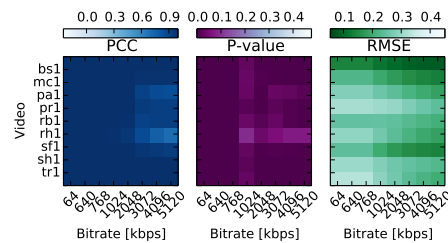
A. Compression Experiment

In this first experiment we evaluated our UDL-based method on video that had been distorted only through compression. This test focused on the 72 impaired videos of the data set streamed at 0% packet loss.

Columns 2 to 4 of Table I depicts the correlation values for all the video types (rows 1 to 9), including the results aggregated over all videos (last row). Overall, the UDL-based method achieves an averaged (last row) PCC correlation (2nd column) higher than 90% and a very low P-value (denoting that our method statistically belongs to the same distribution of VQM), while maintaining and RMSE (4th column) lower than 17%. If we now have a look at the colormaps (Figure 2a), overall, these are dominated by dark colors, demonstrating the high level of accuracy of our method.
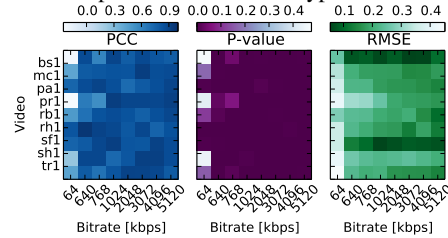
B. Network Loss Experiment

In the second experiment, we extend the analysis to the whole data set (videos are impaired both by compression and network packet loss). We use the same RBM model, trained on the 9 original unimpaired videos, to evaluate the whole data set of compression and network-impaired videos.

The last three columns (5 to 7) of Table I show the correlation results (PCC, P-value and RMSE) per video type and the average across the whole data set. The average PCC (fifth column) is maintained above 75% with a very low deviation. The P-value is below 6%, while the RMSE is kept under 20%.

(a) Compression Experiment: Per video type and bitrate (PLR 0%)

(b) Network Loss Experiment: Per video type, bitrate and packet loss (PLR 0%-10%)

Fig. 2: Accuracy of our method by means of PCC (first column), P-value (second column) and RMSE (third column) to the benchmark quality (VQM). Darker colors denote higher levels of correlation to the benchmark. White represents anti-correlation to the benchmark

Now, if we look at the colormaps (Figure 2b), our method provides very high levels of accuracy for nearly all the video types and bitrate levels. The performance drops slightly for the lowest bit rate variant (64kbps). This was to be expected, since very low bit rate videos (which are nevertheless hardly used nowadays) can suffer from unpredictable behavior deriving from a combination of high compression and packet loss.

V. Conclusion

Accuracy, real-timeliness, adaptability and scalability are all fundamental requisites for a satisfactory video quality assessment of video streaming services. In this letter, we have presented an unsupervised deep learning based-method for online video quality assessment. To our knowledge this is the first to fulfill all these characteristics. Accuracy has been shown using the representative LIMP Video Quality Database (a network impaired video-set consisting of 864 videos) [22], achieving on average between 78% and 91% correlation with VQM. Adaptability is fulfilled by the unsupervised nature of our approach that, unlike supervised learning solutions, does not require labeled training data. Therefore, it much faster and more easily adapts its model as new videos are added. Finally, the scalability of our approach has also been demonstrated in our experimental analysis, where only the 9 original video type samples are sufficient to accurately assess the remaining 864 videos of the dataset. Our approach has significant applicability potential, particularly in the context of adaptive content delivery and quality of experience management in networks.

References

[1] L. Atzori, C. Chen, T. Dagiuklas, and H. R. Wu, "QoE management in emerging multimedia services," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 18–19, 2012.

[2] Y. Fang, W. Lin, and S. Winkler, "Review of existing objective QoE methodologies," in *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*, T. Dagiuklas, C. Chen, P. Chatzimisios, and L. Atzori, Eds. Wiley, 2016.

[3] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison." *IEEE Transactions on Broadcasting*, no. 2, pp. 165–182.

[4] P. Paudyal, F. Battisti, and M. Carli, "Impact of video content and transmission impairments on quality of experience," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 16 461–16 485, 2016.

[5] L. Atzori, A. Floris, G. Ginesu, and D. D. Giusto, "Quality perception when streaming video on tablet devices," *J. Visual Communication and Image Representation*, vol. 25, no. 3, pp. 586–595, 2014.

[6] A. Kordelas, I. Politis, and T. Dagiuklas, "Transport analysis and quality evaluation of MVC video streaming," *Multimedia Tools Appl.*, vol. 75, no. 10, pp. 5619–5644, 2016.

[7] M. Shahid, A. Rossholm, B. Lövström, and H. Zepernick, "No-reference image and video quality assessment: a classification and review of recent approaches," *EURASIP J. Im. and Vid. Proc.*, vol. 2014, p. 40, 2014.

[8] Q. Huynh-Thu and M. Ghanbari, "No-reference temporal quality metric for video impaired by frame freezing artefacts," in *Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt*, 2009, pp. 2221–2224.

[9] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the quality of experience of http video streaming," in *Integrated Network Management*, N. Agoulmine, C. Bartolini, T. Pfeifer, and D. O'Sullivan, Eds. IEEE, 2011, pp. 485–492.

[10] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 207–210, 2015.

[11] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference video quality assessment of compressed video sequences," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 10, pp. 1441–1456, 2012.

[12] K. Zeng and Z. Wang, "Temporal motion smoothness measurement for reduced-reference video quality assessment," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010, 14-19 March 2010, Sheraton Dallas Hotel, Dallas, Texas, USA*, 2010, pp. 1010–1013.

[13] S. Wolf and M. H. Pinson, "Low bandwidth reduced reference video quality monitoring system," in *in Proc. Int. Workshop Video Process. Quality Metrics Consumer Electr.*, Jan. 2005.

[14] S. Yang, "reduced reference mpeg-2 picture quality measure based on ratio of dct coefficients," *Electronic letters*, vol. 47, no. 6, 2011.

[15] J. P. Vink and G. de Haan, "No-reference metric design with machine learning for local video compression artifact level," *J. Sel. Topics Signal Processing*, vol. 5, no. 2, pp. 297–308, 2011.

[16] M. Shahid, J. Panasiuk, G. V. Wallendael, M. Barkowsky, and B. Lövström, "Predicting full-reference video quality measures using HEVC bitstream-based no-reference features," in *Seventh International Workshop on Quality of Multimedia Experience, QoMEX 2015, Pilos, Messinia, Greece, May 26-29, 2015*, 2015, pp. 1–2.

[17] Y. Xue, B. Erkin, and Y. Wang, "A novel no-reference video quality metric for evaluating temporal jerkiness due to frame freezing," *CoRR*, vol. abs/1411.1705, 2014.

[18] T. Shanableh, "A regression-based framework for estimating the objective quality of HEVC coding units and video frames," *Signal Processing: Image Communication*, vol. 34, pp. 22–31, 2015.

[19] F. J. Suárez, A. García, J. C. Granda, D. F. García, and P. Nuño, "Assessing the qoe in video services over lossy networks," *J. Netw. Syst. Manage.*, vol. 24, no. 1, pp. 116–139, Jan. 2016.

[20] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," in *Parallel Distributed Processing: Volume 1: Foundations*, D. E. Rumelhart, J. L. McClelland *et al.*, Eds. Cambridge: MIT Press, 1987, pp. 194–281.

[21] D. C. Mocanu, G. Exarchakos, and A. Liotta, "Deep learning for objective quality assessment of 3d images," in *Image Processing (ICIP), 2014 IEEE International Conference on*, Oct 2014, pp. 758–762.

[22] M. Torres Vega and A. Liotta, "LIMP-Video Quality Database," https://www.tue.nl/index.php?id=53688.

[23] M. Torres, V. Sguazzo, D. C. Mocanu, and A. Liotta, "An experimental survey of no-reference video quality assessment methods," *Int. J. Pervasive Computing and Communications*, vol. 12, no. 1, pp. 66–86, 2016.

[24] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[25] D. Mocanu, G. Exarchakos, H. Ammar, and A. Liotta, "Reduced reference image quality assessment via boltzmann machines," in *Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on*, May 2015, pp. 1278–1281.

[26] Y. Bengio, "Learning deep architectures for ai," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, Jan. 2009.

[27] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," http://www.deeplearningbook.org, 2016, book in preparation for MIT Press.

[28] H. B. Ammar, E. Eaton, M. E. Taylor, D. C. Mocanu, K. Driessens, G. Weiss, and K. Tuyls, "An automated measure of mdp similarity for transfer in reinforcement learning," in *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[29] M. G. Kendall, A. Stuart, and J. K. Ord, Eds., *Kendall's Advanced Theory of Statistics*. New York, NY, USA: Oxford University Press, Inc., 1987.

[30] T. Tieleman, "Training restricted boltzmann machines using approximations to the likelihood gradient," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 1064–1071.

[31] G. Desjardins, A. Courville, Y. Bengio, P. Vincent, and O. Delalleau, "Tempered Markov Chain Monte Carlo for training of restricted Boltzmann machines," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, May 13-15, 2010, Chia Laguna Resort, Sardinia, Italy*, Y. W. Teh and M. Titterington, Eds., 2010, pp. 145–152.

[32] T. Tieleman and G. Hinton, "Using fast weights to improve persistent contrastive divergence," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 1033–1040.

[33] G. E. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.

[34] G. Hinton, "A practical guide to training restricted boltzmann machines," in *Neural Networks: Tricks of the Trade*, ser. Lecture Notes in Computer Science, G. Montavon, G. Orr, and K.-R. Mller, Eds. Springer Berlin Heidelberg, 2012, vol. 7700, pp. 599–619.

[35] D. Mocanu, A. Liotta, A. Ricci, M. Vega, and G. Exarchakos, "When does lower bitrate give higher quality in modern video services?" in *Network Operations and Management Symposium (NOMS), 2014 IEEE*, May 2014, pp. 1–5.

[36] A. Liotta, D. C. Mocanu, V. Menkovski, L. Cagnetta, and G. Exarchakos, "Instantaneous video quality assessment for lightweight devices," in *Proc. of Int. Conference on Advances in Mobile Computing*, ser. MoMM '13, New York, NY, USA, 2013, pp. 525:525–525:531.

[37] C. Perra, "A low Computational Complexity Blockiness Estimation Based on Spatial Analysis," in *in IEEE 22nd Telecommunications Forum*, 2014.

[38] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *Signal Processing Letters, IEEE*, vol. 4, no. 11, pp. 317–320, Nov. 1997.

[39] M. G. Choi, J. H. Jung, and J. W. Jeon, "No-reference image quality assessment using blur and noise," *Int. J. of Elect., Comp., Ener., Electronic and Comm. Eng.*, vol. 3, no. 2, pp. 184 – 188, 2009.

[40] S. Borer, "A model of jerkiness for temporal impairments in video transmission," in *in proc. of the Second International Workshop on Quality of Media Experience (QoMEX)*, 2010.

[41] M. Torres Vega, E. Giordano, D. C. Mocanu, and A. Liotta, "Cognitive no-reference video quality assessment for mobile streaming services," in *in proc. of the 7th International Workshop on Quality of Multimedia Experience (QoMex)*, May 2015.

[42] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *Trans. Img. Proc.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[43] PacketStorm, "Packetstorm hurricane ii network emulator," Available at http://packetstorm.com/packetstorm-products/hurricane-ii-software/.