Digital Object Identifier 10.1109/OJSP.2021.3105305

Editorial
Applied Artificial Intelligence and Machine Learning for Video Coding and Streaming

We are pleased to present this Special Issue on this timely and impactful topic. In the last few years, hardly a day goes by that we do not hear about the latest advancements and improvements that Artificial Intelligence (AI), and particularly its subset Machine Learning (ML) more recently, have brought to a wide spectrum of domains: from technology and medicine to science and sociology, and many others. AI is one of the core enabling components of the fourth industrial revolution that we are currently witnessing, and the applications of AI are truly transforming our world and impacting all facets of society, economy, living, working, and technology. The field of Video Technology is no exception and has already been impacted by Applied AI. Video continues to be the dominant traffic on the Internet, and especially due to the COVID-19 pandemic causing increased video usage, video traffic in the USA increased by 70% in 2020 compared to 2019 and constituted 71% of all 2020 IP traffic.[1] Therefore, improving video coding methods and video networking schemes is vital to cope with this increasing demand. In recent years, we have witnessed an exponential growth of applying AI to revolutionize the field of video coding and streaming. This is the result of AI and ML having become affordable and practical partially due to accessibility to high processing power, GPUs, and availability of various large datasets.

AI and ML-based solutions now offer state-of-the-art in many high-level and low-level image and video related tasks, such as object detection, tracking, segmentation, denoising, filtering, color correction, etc. The power of these tools has recently been introduced to some video coding and video streaming problems as well. A range of Convolutional Neural Network (CNN)-based video coding tools (rate-distortion optimization, deblocking filters, interpolation filters, chroma from luma prediction methods), learned entropy coding, end-to-end image compression techniques, decision tree based encoder speed ups, Fuzzy network bandwidth prediction, and video network resource allocation via reinforcement learning are among these efforts. However, much more effort is required to advance this field, and overcome the existing challenges. Improving the compression efficiency, lowering the overhead of computations of these AI/ML tools, finding suitable loss functions and optimization algorithms, managing network resources according to user experience, accurately predicting network status, and working towards explainable AI are some of these challenges. This Special Issue is dedicated to the said challenges and covers novel applied methods, designs, and systems for AI and ML-based video coding and streaming. Specifically, we invited papers that cover the recent contributions of ML to advances in the whole video processing chain - from creation of high quality video, video compression and flexible representation for optimised distribution, to evaluation of the final Quality of Experience (QoE), all more efficiently achieved thanks to application of ML.

It is interesting to note that 7 of the 8 accepted papers either come from the industry or are the result of academia and industry collaboration. We are particularly pleased to see the participation of industry heavyweights such as (in alphabetic order) AVIWEST, BBC, Bitmovin, Ciena, Google, Netflix, and Qualcomm. This is different from the usual academia-dominated special issues of signal processing journals, most likely because of the issue's focus on applied signal processing and ML, and indicates the immediate and real impact that ML is already having in this field. With this in mind, let us now take a look at the summary of the accepted papers and their topics.

Reflecting the trend of higher video quality demand, powerful deep learning methods based on CNNs have recently been successfully deployed in tasks that aim at making pixels better during content production stages. A significant application area of such approaches is enhancement of legacy content which is available in lower qualities despite being professionally produced. Addressing the need for creation of higher frame rate video, the first paper in this special issue proposes "PDWN: Pyramid Deformable Warping Network for Video Interpolation," Chen *et al.* [A1]. This approach improves existing methods for frame interpolation in multiple aspects thanks to the coarse-to-fine successive refinement approach with deformable convolutions and feature correlations, which also results in smaller model size and shorter inference time. Interestingly, the model can be extended to use more than the typical two frames, which opens various possibilities for new applications of video enhancement.

While new professionally produced content increases user expectations of video quality, the vast amount of complex and diverse User Generated Content (UGC) is created without much quality control. However, users are consuming an increasing amount of UGC, creating the need for fast quality prediction of such content. This challenge is addressed in a paper titled "RAPIQUE: Rapid and Accurate Video Quality Prediction of User Generated Content," Tu *et al.* [A2]. Because of the diversity of UGC, its quality prediction, even if limited to the spatial domain only, is very complex, while available methodologies for evaluating temporal video aspects

[1] Dean Takahashi, "Comcast: Pandemic drove peak internet traffic up 32% in 2020", VentureBeat, March 2, 2021.

are limited. Therefore, this paper introduces a carefully designed quality evaluation framework based on the newly introduced temporal statistics model and an efficient spatial feature extraction model, enabling efficient prediction of UGC video quality.

To enable mass scale content distribution, it is essential to deploy efficient compression techniques. A very high compression significantly reduces video quality, resulting in a poor QoE for the end users. To prevent this, originally high quality videos have to be compressed in multiple representations to adapt to the best conditions (e.g. resolution, bandwidth) available to individual users, using compression schemes that offer the highest quality of video for given conditions. Traditionally designed with only slight help of ML, highly optimised compression solutions are nowadays being revisited and further improved thanks to advances in deep learning. Two distinctive approaches are actively being investigated: more experimental frameworks which are based on ML in an end-to-end fashion, and those that enhance conventional compression frameworks thanks to effective learned optimisations.

To bring end-to-end deep learning methods closer to practical application scenarios, the paper "Transform Network Architectures for Deep Learning based End-to-End Image/Video Coding in Subsampled Color Spaces," Egilmez *et al.* [A3] proposes designs that efficiently handle the most commonly used YUV 4:2:0 video format. In addition to direct extensions of existing approaches for RGB sequences adapted to YUV 4:2:0, this paper provides analysis of tailored extensions, resulting in a proposed framework which increases compression gain and reduces computational requirements. The applicability potential of the proposed approach is evaluated in comparison with test models of HEVC and VVC standards.

The paper "Improved CNN-based learning of interpolation filters for low-complexity inter prediction in video coding," Murn *et al.* [A4] demonstrates how deep learning can help in the design of better pixel prediction within a conventional compression pipeline. In particular it shows how motion compensation can be improved thanks to incorporation of learned interpolation filters within the current state-of-the-art Versatile Video Coding (VVC) framework. A new CNN architecture is used to learn improved interpolation using novel training methodologies. Furthermore, the paper shows that deep networks can be simplified to reduce complexity at the inference stage, without affecting the prediction performance. Such an approach demonstrates potential for improving compression performance compared to the VVC.

As in many applications it is necessary to reduce compressed video bit rate to the levels that introduce compression artefacts, CNN-based approaches can help in restoring some of the original video quality. "A CNN-based Prediction-Aware Quality Enhancement Framework for VVC," Nasiri *et al.* [A5] paper introduces a framework that models CNN enhancements taking into account parameters of the underlying coding scheme (in this case the VVC). In this way, the quality enhancement is tailored to specific compression artefacts, which then leads to further improvement of video quality. Presented approaches address both in-loop filtering (normative) and post-processing. While the post processing can be regarded as the process that takes place at the user side, it can also be used before transcoding. Therefore, this paper contributes to both the video quality enhancement and video compression pipelines.

Another paper, titled "Fast Multi-Resolution and Multi-Rate Encoding for HTTP Adaptive Streaming Using Machine Learning," Çetinkaya *et al.* [A6] proposes an ML-based approach for fast encoding of multiple video representation using HEVC, which are needed for adaptive streaming. To avoid wide search of highly optimised parameters for each required rate and resolution representation of a video, the presented approach reuses encoding information from selected reference representations. Penalties on compression efficiency using such an approach are very low, while significant time savings are achieved.

To further optimise video streaming, a number of different video representations can be reduced based on the underlying content itself. However, selection of such a limited number of content-dependent representations may still require creation of a number of representations, which is an unnecessarily complex process. Paper "Efficient Bitrate Ladder Construction for Content-Optimised Adaptive Video Streaming," Katsenou *et al.* [A7] addresses this problem by proposing a solution which selects required video representations within a defined bitrate range. Specifically, the paper takes into account spatial and temporal content features which enable accurate selection of content representations, significantly reducing the costs of previously reported approaches.

Finally, in "Forecasting Video QoE with Deep Learning from Multivariate Time-series," Dinaki *et al.* [A8] authors propose a hybrid BiLSTM-CNN model to forecast the near-future value of video QoE, unlike existing works that estimate or predict the current value of video QoE. This enables the video service provider to prognosticate near-future low QoE occurrences, before they actually happen, and take corrective actions to prevent those occurrences and to deliver a smoother experience to video consumers. Such a prognostication system can be used for fault diagnosis in packet video networks, anomaly detection in CDN and wireless networks, and resource scheduling in wireless networks for adaptive DASH delivery.

In the end, we would like to thank all the authors who submitted their papers, the reviewers for generously giving their time and expertise, OJSP EiC Prof. Mari Ostendorf and OJSP Administrator Ms. Rebecca Wollman for their help and support during the submission and reviewing process.

MARTA MRAK, *Guest Editor*
BBC R&D
London, U.K.

MAHMOUD REZA HASHEMI, *Guest Editor*
University of Tehran
Tehran, Iran

SHERVIN SHIRMOHAMMADI, *Guest Editor*
University of Ottawa
Ottawa, Canada

YING CHEN, *Guest Editor*
Alibaba Cloud Intelligence Group
Hangzhou, China

MONCEF GABBOUJ, *Guest Editor*
Tampere University
Tampere, Finland

APPENDIX: RELATED ARTICLES

[A1] Z. Chen, R. Wang, H. Liu, and Y. Wang, "PDWN: Pyramid deformable warping network for video interpolation," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3075879.

[A2] Z. Tu *et al.*, "RAPIQUE: Rapid and accurate video quality prediction of user generated content," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3090333.

[A3] H. E. Egilmez *et al.*, "Transform network architectures for deep learning based end-to-end image/video coding in subsampled color spaces," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3092257.

[A4] L. Murn, S. Blasi, A. F. Smeaton, and M. Mrak, "Improved CNN-based learning of interpolation filters for low-complexity inter prediction in video coding," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3089439.

[A5] F. Nasiri, W. Hamidouche, L. Morin, N. Dhollande, and G. Cocherel, "A CNN-based prediction-aware quality enhancement framework for VVC," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3092598.

[A6] E. Çetinkaya, H. Amirpour, C. Timmerer, and M. Ghanbari, "Fast multi-resolution and multi-rate encoding for HTTP adaptive streaming using machine learning," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3078657.

[A7] A. V. Katsenou, J. Sole, and D. R. Bull, "Efficient bitrate ladder construction for content-optimized adaptive video streaming," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3086691.

[A8] H. E. Dinaki, S. Shirmohammadi, E. Janulewicz, and D. Côté, "Forecasting video QoE with deep learning from multivariate time-series," *IEEE Open J. Signal Process.*, vol. 2, 2021, doi: 10.1109/OJSP.2021.3099065.