Guest Editorial
Introduction to Special Section on Learning-Based Image and Video Compression

VIDEO is being watched more than ever before. It is estimated that in 2020, 82% of global IP traffic and 79% of global Internet traffic will come from video; globally 3 trillion minutes (5 million years) of video content will cross the Internet each month. According to the Cisco 2020 Forecast, that is one million minutes of video streamed or downloaded every second [1]. The rapidly increasing consumption of storage capacity and transmission bandwidth from video, especially HD and UHD video content, has made video compression a critical stage to guarantee the quality of delivery and playback.

Historically, video as well as image was stored as an analog signal. With the invention of DCT in the 1970s, researchers and engineers started experimenting with DCT for digital image and video compression, which led to the development of H.261, and JPEG. H.261 is generally considereds the first practical video coding standard, which consists of block-based inter picture prediction with motion compensation, transform and entropy coding. The mainstream video coding standards developed since then, such as MPEG-2, H.263, H.264/AVS, H.265/HEVC and the emerging VVC, as well as many widely industry adopted video compression codecs may be seen as expansions from this block-based structure. As of today, many block-based video coding technologies have been adopted into international standards and used to enhance video compression quality, such as intra prediction, in-loop filtering, more comprehensive transform, entropy coding, inter picture prediction and more flexible block partitioning, etc. Together they have increased the compression ratio by approximately 10 to 20 times compared with the first-generation video coding standard H.261. Despite many technical advancements over the years, the focus of the block-based hybrid coding structure remains to be improving the compression-complexity performance of each coding module, as well as the interaction among related modules, to achieve overall compression quality enhancement.

Deep learning has demonstrated its superior capability for solving computer vision and image processing problems in the last decade. Witnessing such success, researchers and engineers are motivated to investigate learning-based technologies for image and video compression. In fact, some researchers started exploring the utilization of neural networks for image compression as early as in the 1990s. However, the computing power back then could not afford training and solving complex models. Hence, the compression results did not seem very promising. In recent years, with the advent of greater and cheaper computing power, and a huge amount of training data, learning-based video and image coding tools have regained a lot of research interest. Some encouraging progress and evidence have been demonstrated in the last five years. These research works can be divided into two categories: end-to-end learning-based coding schemes, and learning-based coding tools that are embedded into conventional coding schemes, such as the block-based hybrid coding schemes.

The first category completely gets rid of the conventional hybrid coding structure and compression is performed through an end-to-end learning scheme, usually a deep neural network. The interest in this category has rapidly grown since Google presented a general framework for image compression using recurrent networks in 2015. After five years of continuous research and experiments, for image coding and all-intra video coding, some end-to-end learning-based schemes have reportedly achieved compression efficiency comparable to the state-of-the-art block-based hybrid coding schemes, e.g. HEVC intra or the emerging VVC intra. The technology advances and evidence from academic research also caught the attention of the industry. In January 2020, JPEG issued a Call for Evidence (CfE) on Learning-based Image Coding Technologies; in April 2020, the Future Video Coding Study Group (FVC SG) of IEEE Data Compression Standard Committee issued another CfE on Deep Learning-Based Image Compression with some different emphases and considerations. Deep neural network-based end-to-end image and video coding is also being investigated in Audio and Video Coding Standard Working Group (AVS).

The second category is a mixture of conventional block-based hybrid coding structure and learning-based methods, wherein the learning-based coding tools are integrated into the block-based system, either as an independent module, e.g. a learning-based in-loop filter, or as part of a coding stage, e.g. one of the intra prediction modes. While the first category has shown compression benefits on images and all-intra coded videos, the second category has reported coding gains on top of block-based hybrid video coding, such as HEVC. In 2017, ISO/IEC MPEG and ITU-T VCEG issued a joint Call for Proposals (CfP) on video compression with capability beyond HEVC. In early 2018, a total of 46 responses from 33 global research and industrial organizations were received and evaluated at the 122$^{\text{nd}}$ MPEG meeting. Among the received CfP responses and technical proposals, about 10 included learning-based tools which reported compression

Digital Object Identifier 10.1109/TCSVT.2020.2995955

benefits against or on top of conventional methods. These tools and solutions are introduced in detail in the TCSVT special section article "Deep learning-based technology in responses to the joint call for proposals on video compression with capability beyond HEVC" by D. Liu *et al.* An ad-hoc group (AHG9) was set up for investigating the compression performance and complexity of neural network-based coding tools and an output document JVET-M1006 was issued in January 2019 to provide thorough evaluation results as well as evaluation matrix and methodology. With VVC being finalized, MPEG established another ad-hoc group to continue investigating the performance improvement potential of deep neural network-based video coding for both hybrid and end-to-end coding system in April 2020. During the exploration for the next-generation video coding tools beyond AV1 conducted by the Alliance of Open Media (AOM), several learning-based coding tools have been proposed and are under consideration. Learning-based image and video compression are also studied in AVS under the context of AVS3 since 2017, where CNN-based in-loop filtering, intra and inter prediction tools have been proposed to replace or supplement conventional coding methods and demonstrated additional coding gains.

This Special Section consists of 12 articles with one emphasizing end-to-end learning-based image compression, and the rest focusing on learning-based coding tools. The learning-based coding tool articles are further split into five sub-categories, i.e., learning-based coding tools for intra prediction, inter prediction, filtering, arithmetic coding, and encoder optimization.

The first article "Toward variable-rate generative compression by reducing the channel redundancy" by C. Han *et al.* presents an end-to-end autoencoder for image compression. Based on the training framework of generative adversarial networks, the autoencoder is trained to minimize a reconstruction loss and an adversarial loss. Notably, the encoder part of the autoencoder is regularized by a rate loss that aims at minimizing the feature variance across channels, and thus allowing variable rate compression to be achieved at inference time with a single model by spatially variant feature masking.

The subcategory of learning-based intra prediction has three articles touching upon H.265/HEVC intra-picture prediction and lossless image coding. "Multi-scale convolutional neural network-based intra prediction for video coding" by Y. Wang *et al.* employs two CNNs to improve the angular intra prediction in H.265/HEVC, wherein the first extracting multiscale features from the angular prediction signal along with pixels in an L-shaped causal neighborhood, and the second decoding these features into a better-quality predictor. The pixels in the causal neighborhood serve as contextual information to facilitate the decoding process. Along this line of research, in their article "CNN-based intra-prediction for lossless HEVC," I. Schiopu *et al.* adapt the structure of the causal neighborhood according to the selected angular mode. Departing from block-wise intra prediction, another article "Deep-learning-based lossless image coding" by I. Schiopu *et al.* forms a CNN-based prediction of residual signals resulting from pixel-wise intra prediction.

Two articles have been selected to demonstrate research progress of CNN-based inter-picture prediction. "Deep frame prediction for video coding" by H. Choi *et al.* presents a new frame prediction method using jointly or separately trained deep CNNs for uni-directional and bi-directional predictions. The prediction result is used as an additional inter prediction mode for which no motion vectors are involved. "Convolutional neural network based bi-prediction utilizing spatial and temporal information in video coding" by J. Mao *et al.* feeds spatial neighboring pixels and temporal distances into the proposed CNN model in additional to reference and prediction blocks to utilize temporal and spatial correlation of pixels simultaneously and harmonize interpolation and extrapolation prediction. Deep neural network helps to refine traditional bi-hypothesis AMVP and merge/skip modes.

Among all coding modules or stages in the hybrid system, in-loop filtering is probably the one which has caught the most research interest with shown potential to improve compression efficiency, both objectively and subjectively, by using neural networks. "A switchable deep learning approach for in-loop filtering in video coding" by D. Ding *et al.* introduces a so-called Squeeze-and-Excitation Filtering CNN (SEFCNN) as an optional in-loop filter in addition to conventional in-loop filters. The proposed SEFCNN is further comprised of two subnets to capture the non-linear interaction between channels. "Recursive residual convolutional neural network-based in-loop filtering for intra frames" by S. Zhang *et al.* proposes a recursive residual convolution neural network (RRCNN)-based in-loop filters to improve the quality of reconstructed intra frames while reducing the bitrate at the same time. In contrast to the previous article, where different models are trained for high and low bit rates, a single model is used to handle various bit rates in this article, with different networks designed for filtering luma and chroma components. In both articles, a switchable mechanism is adopted to allow the system to choose between conventional in-loop filters, e.g. HEVC deblocking filter and SAO, and the proposed CNN-based in-loop filter at frame, region or CTU level based on the rate-distortion cost.

Instead of using CNNs for reconstruction or prediction of pixel values, the article "Convolutional neural network-based arithmetic coding for HEVC intra-predicted residues" by C. Ma *et al.* adapts their use for estimating the probabilities of intra-predicted residues to be encoded with arithmetic coding. Specifically, the probability of a transform coefficient associated with intra-predicted residual is estimated by a CNN that takes as input the reconstruction of neighboring blocks and that of the current block using its previously encoded transform coefficients.

Learning techniques also find applications in fast mode decision and perceptual coding. The article "DeepSCC: Deep learning-based fast prediction network for screen content coding" by W. Kuang *et al.* exploits hierarchical deep features extracted via layers of convolution to predict the mode candidates for coding units at all depth levels of the quadtree partitioning for screen content coding. In contrast, "Fast depth map intra coding for 3D video compression-based tensor feature extraction and data analysis" by H. Hamout *et al.* compares hand-crafted tensor features of a coding unit against threshold values derived from automatic merging possibilistic clustering, to reduce its mode set for fast depth map intra coding with 3D-HEVC. The last article "High-definition video compression system based on perception guidance of salient information of a convolutional neural network and HEVC compression domain" by S. Zhu *et al.* presents perceptual video coding, where they adapt the rate-distortion optimization and the selection of quantization parameter in HEVC according to a spatiotemporal saliency map computed from CNN-based spatial saliency prediction and temporal saliency estimated by block-based motion vectors.

Even though video and image coding have been a research area for a few decades, the interest from academia and industry has never faded. On the other side, deep learning and AI have made great progress in recent years and will continue to be hot research subjects. With huge market demand and evidence of some learning-based video and image processing tools already having landed in real-world products, there is no reason for us not to expect a boom in learning-based coding technologies soon, if not already. Due to limited space, this special section could only catch a small portion of related research work. We hope you enjoy reading it and find it helpful to your research on learning-based video and image coding technologies.

SHAN LIU, *Lead Guest Editor*
Tencent America
Palo Alto, CA 94306 USA
e-mail: shanl@tencent.com

WEN-HSIAO PENG, *Guest Editor*
Department of Computer Science
National Chiao Tung University
Hsinchu 30010, Taiwan
e-mail: wpeng@cs.nctu.edu.tw

LU YU, *Guest Editor*
Department of Information and
Communication Engineering
Zhejiang University
Hangzhou 310027, China
e-mail: yul@zju.edu.cn

REFERENCES

[1] *CISCO Global—Forecast Highlights*, 2020, pp. 2–3.

Shan Liu (Senior Member, IEEE) received the B.Eng. degree in electronic engineering from Tsinghua University, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California.

She was the Director of the Media Technology Division, MediaTek USA. She was with MERL, Sony, and IBM. She has been actively contributing to international standards over the last decade. She is currently a Tencent Distinguished Scientist and the General Manager of Tencent Media Lab. She has served as a Co-Editor for HEVC SCC and the emerging VVC. She had numerous technical contributions adopted into various standards, such as HEVC, VVC, OMAF, DASH, and PCC. At the same time, technologies and products developed by her and under her leadership are serving over 100 million daily active users. She has published more than 80 journal and conference papers. She holds more than 150 granted U.S. and global patents. Her research interests include audiovisual, high volume, immersive and emerging media compression, intelligence, transport, and systems. She served on the Industrial Relations Committee of the IEEE Signal Processing Society from 2014 to 2015. She is serving on the Editorial Board of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY for the period of 2018–2021. She was the Vice President of the Industrial Relations and Development of the Asia–Pacific Signal and Information Processing Association from 2016 to 2017. She was named the APSIPA Industrial Distinguished Leader in 2018. She was appointed as the Vice Chair for the IEEE Data Compression Standards Committee in 2019.

Wen-Hsiao Peng (Senior Member, IEEE) received the Ph.D. degree from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 2005. He was with the Intel Microprocessor Research Laboratory, Santa Clara, CA, USA, from 2000 to 2001, where he was involved in the development of the International Organization for Standardization (ISO) Moving Picture Experts Group (MPEG)-4 fine granularity scalability and demonstrated its application in 3D peer-to-peer video conferencing. Since 2003, he has actively participated in the ISO MPEG digital video coding standardization process and contributed to the development of the High Efficiency Video Coding (HEVC) standard and MPEG-4 Part 10 Advanced Video Coding Amd.3 Scalable Video Coding standard. His research group at NCTU is one of the few university teams around the world that participated in the Call-for-Proposals on HEVC and its Screen Content Coding extensions. He was a Visiting Scholar with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA, from 2015 to 2016. He is currently a Professor with the Computer Science Department, NCTU. He has authored over 70 technical articles in the field of video/image processing and communications and over 60 standards contributions. His research interests include video/image coding, deep/machine learning, multimedia analytics, and computer vision. He is a Technical Committee Member of the Visual Signal Processing and Communications and Multimedia Systems and Application tracks of the IEEE Circuits and Systems Society (CASS). He was the Technical Program Co-Chair of the IEEE VCIP 2011, the IEEE ISPACS 2017, and APSIPA ASC 2018, the Publication Chair of IEEE ICIP 2019, the Area Chair of the IEEE ICME and VCIP, and a Review Committee Member of the IEEE ISCAS. He served as the Associate Editor-in-Chief for *Digital Communications*, a Lead Guest Editor/Guest Editor/SEB Member for the IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS, an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and a Guest Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS. He was an elected a Distinguished Lecturer of APSIPA and the Chair-Elect of the IEEE CASS VSPC Technical Committee.

Lu Yu (Senior Member, IEEE) received the B.Eng. degree in radio engineering and the Ph.D. degree in communication and electronic systems from Zhejiang University. She is currently a Distinguished Professor with Zhejiang University. She developed a series of theories, algorithms, technologies, and architecture designs of video coding reflected in 150 peer-reviewed articles. She has over 80 granted patents and more than 100 adopted technical contributions that help to define a series of IEEE, ISO/IEC, as well as ISO/IEC and ITU-T joint standards, including IEEE 1857 (AVS1), IEEE 1857.4 (AVS2), AVC, HEVC, and MPEG-B—Part 2. She was the Video Subgroup Co-Chair and the Chair of the AVS Working Group from 2002 to 2017. She acts as the Video Subgroup Chair of ISO/IEC JTC1 SC29 WG11, known as MPEG, and leads standardization activities such as immersive video coding, essential video coding, and low-complexity enhancement video coding. She serves on the Editorial Board of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY for the period of 2020–2021 and as the Chair of the Technical Committee of Education and Outreach of the IEEE Society of Circuits and Systems (CASS). She was the General Chair of the Picture Coding Symposium 2019 and the Workshop on Packet Video 2006, the Organization Committee Member of the IEEE Workshop of Multimedia Signal Processing 2011, an Area Editor of the *Signal Processing: Image Communication* (EURASIP), and the Chair of the Membership and Election Subcommittee of Technical Committee of Visual Signal Processing and Communication, CASS.