Grand Challenge on Neural Network-based Video Coding (ISCAS 2024)

Abstract

Recently, there is increasing interest in neural network-based video coding, including end-to-end and hybrid schemes. To foster the research in this emerging field and provide a benchmark, we propose this Grand Challenge (GC). In this GC, different neural network-based coding schemes will be evaluated according to their coding efficiency and innovations in methodologies. Three tracks will be evaluated, including the 1) hybrid neural network-based (NN-based) video codec, 2) end-to-end video codec, and 3) neural network enhanced VVC encoder. In the hybrid codec track, deep network-based coding tools shall be used with traditional video coding schemes. In the end-to-end codec track, the whole video codec system shall be built primarily upon deep networks. In the neural network enhanced VVC encoder track, deep network-based encoding algorithms can be applied in a VVC encoder which generates VVC compatible bitstreams.

Participants shall express their interest to participate in this Grand Challenge by sending an email to the organizer Dr. Yue Li and are invited to submit their schemes as ISCAS papers. The papers will be regularly reviewed and, if accepted, must be presented at ISCAS 2024. The submission instructions for Grand Challenge papers will be communicated by the organizers.

Rationale

In recent years, deep learning-based image/video coding schemes have achieved remarkable progress. As two representative approaches aiming at future video codec schemes, hybrid solutions and end-to-end solutions have both been investigated extensively. Hybrid solutions adopt deep network-based coding tools to enhance traditional video coding schemes while end-to-end solutions build the whole compression scheme based on deep networks. Besides, NN-based methods are also widely studied to optimize or speed up encoders compliant to existing popular standards such as VVC. Although great advancement has been observed, there are still numerous challenges remaining to be addressed:

- How to harmonize a deep coding tool with a hybrid video codec, for example, how to take compression process into consideration when developing a deep tool for pre-processing;
- How to exploit long-term temporal dependency in an end-to-end framework for video coding;
- How to leverage automated machine learning-based network architecture optimization for higher coding efficiency;
- How to perform efficient bit allocation with deep learning frameworks;

- How to achieve a better global result in terms of rate-distortion trade-offs, for example, to take the impact of the current step on later frames into account, possibly by using reinforcement learning;
- How to achieve better complexity-efficiency trade-offs;
- How to speed up a VVC encoder with less coding efficiency loss via NN methods or use NN-based preprocessing to enhance the VVC encoding efficiency.

In view of these challenges, several activities towards improving deep-learning-based image/video coding schemes have been initiated. For example, a special section on "Learning-based Image and Video Compression" was published in TCSVT, July 2020; a special section on "Optimized Image/Video Coding Based on Deep Learning" was published in OJCAS, December 2021; and the "Challenge on Learned Image Compression (CLIC)" at CVPR has been organized annually since 2018. In hopes of encouraging more innovative contributions towards the aforementioned challenges in the ISCAS community, we proposed this grand challenge since 2022. It has been successfully held for two years (ISCAS 2022, ISCAS 2023), attracting related researchers all over the world. As being looked forward by many experts in this area, the grand challenge will be held again for ISCAS 2024, with more tracks and more awards.

Important Dates

20 September 2023: The organizers release the validation set as well as the corresponding test information to those who have expressed interest (for example, frame rates and intra periods) and template for performance reporting (with rate-distortion points for the validation set)

7 November 2023 (extended): Deadline of paper submission (aligned with Special Sessions) for participants

- Submission to this Special Session should be through ePapers: https://epapers2.org/iscas2024

28 December 2023: Participants upload a docker container for the first two tracks (i.e., the hybrid NN-based and end-to-end video codec) with a decoder, wherein only the single decoder shall be utilized for decoding all the bitstreams; or for the third track (i.e., NN enhanced VVC encoder track) with an encoder, wherein only the single VVC encoder shall be utilized for generating all the bitstreams.

8 January 2024: Organizers release the test sequences (including frame rate, corresponding rate-distortion points, etc.)

15 January 2024: Paper acceptance notification

31 January 2024: Participants upload compressed bitstreams and decoded YUV files

5 February 2024: Deadline of fact sheets submission for participants

5 February 2024: Camera-ready paper submission deadline

TBA: Paper presentation at ISCAS 2024

TBA: Awards announcement (at the ISCAS 2024 banquet)

Awards

ByteDance will sponsor the awards of this grand challenge. Four categories of awards are expected to be presented. Three top-performance awards will be granted according to the performance, for the hybrid track, the end-to-end track, and the VVC encoder-only track respectively. In addition, to foster innovation, a top-creativity award will be given to the most inspiring scheme recommended by a committee group, and it is only applicable to participants whose papers are accepted by ISCAS 2024. The winner of each award (if any) will receive a $5000 USD prize.

Requirements, Evaluation, Timeline and Awards

- Training Data Set

It is recommended to use the following training data.
UVG dataset: http://ultravideo.cs.tut.fi/
CDVL dataset: https://cdvl.org/
Additional training data are also allowed to be used given that they are described in the submitted document.

- Test Specifications

In the test, each scheme will be evaluated with multiple YUV 4:2:0 test sequences in the resolution of 1920x1080.
There is no constraint on the reference structure. Note that the neural network must be used in the decoding process of the hybrid track and the end-to-end track, while the VVC reference software VTM will be utilized for decoding bitstreams of the NN enhanced VVC encoder-only track.

- Evaluation Criteria

The test sequences will be released according to the timeline and the results will be evaluated with the following criteria:
The decoded sequences will be evaluated in the 4:2:0 color format.
PSNR (6*PSNRY + PSNRU + PSNRV)/8 will be used to evaluate the distortion of the decoded pictures.
Average Bjøntegaard delta bitrates (BDR) [1] for all test sequences will be gathered to compare the coding efficiency.
Anchors of HM 16.22 [2] and VTM-20.2 [3] coded with QPs = {22, 27, 32, 37} under the random access configurations defined in the HM and VTM common test conditions [4, 5] will be provided. Note that the HM anchor is used for the hybrid and end-to-end tracks, while the VTM anchor is used for the VVC encoder-only track. The released anchor data will include the bit-rates corresponding to the four QPs for each sequence.

Additional constraints for the first two tracks (i.e., the hybrid NN-based and end-to-end video codec) are listed as follows:

1. It is required that the proposed method should generate four bit-streams for each sequence, targeting the anchor bit-rates corresponding to the four QPs. For each sequence, the range of four real bit-rates shall be [80% * the lowest anchor bit-rate, 120% * the highest anchor bit-rate];
2. Only one single decoder shall be utilized to decode all the bitstreams;
3. The intra period in the proposed submission shall be no larger than that used by the anchor in compressing the validation and test sequences.

While for the NN enhanced VVC encoder track, the additional requirements are listed as follows:

1. The docker file shall have the capability of encoding the test sequences to generate VTM-compatible bitstreams;
2. It is required that the proposed method should generate four bit-streams for each sequence, targeting at the anchor bit-rates corresponding to the four QPs. For each test point, the bit-rate of the proposed method should be in the range of 90% to 110% of the anchor bit-rate;
3. The VTM-20.2 decoder is utilized to decode generated bitstreams to get reconstructed YUV files and use those YUV files to calculate the PSNR values. All the generated bitstreams MUST be decoded successfully;
4. The VTM-20.2 encoder is utilized as the anchor encoder. For each test point, denote the encoding time of the proposed encoder as T1, the encoding time of VTM-20.2 encoder as T2, T1 and T2 should satisfy: T1 <= 70% T2. Note that T1 and T2 shall be evaluated on the same platform with single thread (e.g., Intel(R) Xeon(R) Platinum 8336C CPU @ 2.30GHz, NVIDIA A100-SXM4-80GB GPU). Encoding time comparison will be verified by the organizers.

Proposed Documents

A docker container with the executable scheme must be submitted for result generation and cross-check. Each participant is invited to submit an ISCAS paper, which must describe the following items in detail.

- The methodology
- The training data set
- Detailed rate-distortion data (comparison with the provided anchor is encouraged)
- Complexity analysis of the proposed solutions is encouraged for the paper submission.

Contact

For any inquiries on this grand challenge, please contact:
Dr. Yue Li yue.li@bytedance.com

References

[1] Bjøntegaard, “Calculation of average PSNR differences between RD-Curves,” ITUT SG16/Q6, Doc. VCEG-M33, Austin, Apr. 2001.
[2] https://vcgit.hhi.fraunhofer.de/jvet/HM/-/tree/HM-16.22
[3] https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-20.2
[4] Common Test Conditions and Software Reference Configurations for HM (JCTVC-L1100)
[5] JVET common test conditions and software reference configurations (JVET-J1010)

Organizer Biographies

Li Zhang received a Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2009. From 2009 to 2011, she held a post-doctoral position at the Institute of Digital Media, Peking University, Beijing. From 2011 to 2018, she was a Senior Staff Engineer at Qualcomm, Inc., San Diego, CA, USA. She is currently the Head of Advanced Video Group, Bytedance Inc., San Diego, CA, USA. Her research interests include 2D/3D image/video coding, video processing, and transmission. She was a Software Coordinator for Audio and Video Coding Standard (AVS) and the 3D extensions of High-Efficiency Video Coding (HEVC). She has authored 450+ standardization contributions, 170+ granted US patents, 90+ technical articles in related book chapters, journals, and proceedings in image/video coding and video processing. She has been an active contributor to the Versatile Video Coding (VVC), Advanced AVS, the IEEE 1857, 3D Video (3DV) coding extensions of H.264/AVC and HEVC, and HEVC screen content coding extensions. During the development of those video coding standards, she co-chaired several ad hoc groups and core experiments. She has been appointed as an Editor of AVS, the Main Editor of the Software Test Model for 3DV Standards. She organized/co-chaired multiple special sessions and grand challenges at various conferences. She is a Senior member of IEEE, serves as associate editors in IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT), Publicity Subcommittee Chair of the Technical Committee member of Visual Signal Processing and Communications in IEEE CAS Society (VSPC TC).

Jizheng Xu received a Ph.D. degree in electrical engineering from Shanghai Jiaotong University, China in 2011. He joined Microsoft Research Asia in 2003 and served as a Research Manager and joined ByteDance multimedia lab as a Research Scientist in 2018. He has authored and co-authored over 140 refereed conference and journal refereed papers. His research interests include image and visual signal representation, image/video compression and communication, computer vision, and deep learning. He has been an active contributor to ISO/MPEG and ITU-T video coding standards, including H.264/AVC, H.265/HEVC, and VVC/H.266. He initiated the screen content coding in H.265/HEVC and was a major technical contributor. He chaired and co-chaired the ad-hoc group of exploration on wavelet video coding in MPEG, and various technical ad-hoc groups in JCT-VC, e.g., on screen content coding, on parsing robustness, on lossless coding. He was an Associate Editor for the IEEE Transactions on Circuits and Systems for Video Technology from 2018 to 2020. He served as a Guest Editor for the special issue on Screen Content Video Coding and Applications of the IEEE Journal on Emerging and Selected Topics in Circuits and

Systems in 2016. He co-organized and co-chaired special sessions on scalable video coding, directional transform, high-quality video coding at various conferences.

Kai Zhang received a B.S. degree in computer science from Nankai University, Tianjin, China, in 2004. In 2011, he received a Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. From 2011 to 2012, he worked as a researcher in Tencent Inc. Beijing, China. From 2012 to 2016, he worked as a technical manager then a team manager in Mediatek Inc. Beijing, China, leading a research team to propose novel technologies to emerging video coding standards. In 2016, he joined Qualcomm Inc. San Diego, CA. Since 2018, he has been doing research work on video coding as a senior research scientist in Bytedance Inc. San Diego, CA. Dr. Zhang' research interests include video/image compression, coding, processing and communication, especially video coding standardization. In 2006, he proposed his first proposal to JVT. From then, he has contributed more than 300 proposals to JVT, VCEG, JCT-VC, JCT-3V, JVET, and AVS team, covering many important aspects of major standards such as H.264/AVC SVC extension, HEVC, 3D-HEVC, VVC and AVS. During the development of VVC, Dr. Zhang co-chaired several core experiments and branch of groups. Currently, Dr. Zhang serves as a coordinator of the reference software known as ECM in JVET, to explore video coding technologies beyond VVC. Dr. Zhang has 300+ granted or pending U.S. patents applications, which are mostly essential to popular video coding standards. Dr. Zhang has authored/co-authored 30+ papers on well-known journals/conferences such as T-IP, T-CSVT, T-B, JETCAS, ICIP, ISCAS, etc. Dr. Zhang is an Associate Editor of IET Image Processing. Dr. Zhang also serves as a reviewer for many journals/conferences.

Yue Li received a B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2014 and 2019, respectively. He is currently a Research Scientist with Bytedance Multimedia Lab, San Diego, CA, USA. His research interests include image/video coding and processing. He has authored 10+ neural network-based standardization contributions to the Versatile Video Coding (VVC). He has authored/co-authored 15+ papers on well-known journals/conferences such as T-IP, T-CSVT, CSUR, ICIP, ICME, DCC, etc. He also serves as a reviewer for those journals/conferences.