Title: Practical end-to-end image/video compression challenge

1. Challenge motivation and description

End-to-end image/video compression has been a research focus for both academia and industry for over six years. A number of technologies have been developed such as expressive auto-encoder neural networks, accurate probability estimation neural networks, and conditional end-to-end video coding framework and so on. Until recently, the performances of both end-to-end image and video compression schemes have surpassed that of the H.266/Versatile Video Coding (VVC) under certain test conditions. To promote its practical use, we think it is time to consider the complexities of the end-to-end image/video compression schemes, especially the decoding complexities.

This challenge calls for novel end-to-end image/video compression algorithms which can result in a good balance between the performance and decoding complexity. Based on our last year's experiences, we will set a proper weight in the quality metric to balance the performance and decoding complexity for both the end-to-end image and video compression tracks. The participates are required to compress all images/videos defined in the Test Dataset. In the end-to-end image compression track, the actual bits per pixel (bpp) is not allowed to exceed a target bpp, which is set to the bpp of the test image coded by BPG using the quantization parameter 28. In the end-to-end video compression track, the actual bitrate is not allowed to exceed a target bitrate (kbps), which is set to the bitrate of the test sequence coded by VTM using the quantization parameter 27 under the random access configuration.

In the following, the detailed information of each track will be further provided.

2. End-to-end image compression track

2.1 Dataset

- Training and Validation Dataset: A collection of about 1600 high-resolution images will be provided as the training and validation dataset. Participates are free to split the provided images into training and validation dataset. Participates are also free to use some other dataset for training and validation.
- Test Dataset: 20 images with resolution 4K will be used for the evaluation. All the images will be in RGB color space and PNG file format. These images will be distributed to all the participates before a certain date. Participates are required to compress them within 72 hours.

2.2 Evaluation metrics

The performance Q will be evaluated by a weighted sum of the delta PSNR and the decoding complexity,

$$Q = w \cdot \Delta PSNR - dTime$$

where PSNR is calculated using the average PSNR of the R, G, and B components. ΔPSNR is calculated by subtracting the PSNR of BPG from that of the proposed method. dTime is measured by the seconds used for both entropy decoding and image reconstruction with a V100 GPU provided by the organizers. Therefore, it is also required for the methods to be decoded successfully by the V100 GPU provided by the organizers. w is set to 1 to achieve a good balance between performance and decoding complexity.

2.3 Submission requirements

- The participates are requested to submit a decoder along with a docker environment and the corresponding script which can run the decoder.
- The participates are requested to submit the compressed bitstreams. The bitstreams shall be named like I01.bin
- The participates are requested to submit the decoded images. The decoded images shall be named like I01dec.png
- The participates can choose to submit a paper describing the end-to-end image compression scheme to the VCIP challenge session or not. Please follow the VCIP paper template to prepare your manuscript if you want to submit a paper.

3. End-to-end video compression track

3.1 Dataset

- Training and Validation Dataset: It is recommended to use the UVG and CDVL dataset for training. Participates are free to split the provided videos into training and validation dataset. Participates are also free to use some other dataset for training and validation.
- Test Dataset: 10 video sequences in the resolution of 1080p will be used for evaluation. Each sequence contains 96 frames. All the sequences will be in YUV 4:2:0 color space. These video sequences will be distributed to all participates before a certain date. Participates are required to compress them within 72 hours.

3.2 Evaluation metrics

The decoded video sequences will be evaluated in YUV 4:2:0 color space. The weighted average $PSNR = (6 \times PSNR_Y + PSNR_U + PSNR_V)/8$ of the Y, U, and V components will be used to evaluate the distortion of the decoded video sequences. An anchor of VTM-17.0 coded with $QP = 27$ under random access configuration defined in the VTM common test conditions (*encoder_randomaccess_vtm.cfg*) will be provided. The actual bitrate (kbps) of the bitstream of each video sequence is not allowed to exceed the target kbps of the test video coded by the anchor. The intra period in the proposed submission shall be no larger than that used by the anchor.

The performance Q will be evaluated by a weighted sum of the delta PSNR and the decoding complexity,

$$Q = w \cdot \Delta PSNR - dTime,$$

where $\Delta PSNR$ is calculated by subtracting the PSNR of VTM from that of the proposed method. $dTime$ is measured by the seconds used for both entropy decoding and video reconstruction with a V100 GPU provided by the organizers. Therefore, it is also required for the methods to be decoded successfully by the V100 GPU provided by the organizers. w is set to 1 to provide a good balance between the complexity and the performance.

3.3 Submission requirements

- The participates are requested to submit a decoder along with a docker environment and the corresponding script which can run the decoder.
- The participates are requested to submit the compressed bitstreams. The bitstreams shall be named like V01.bin
- The participates are requested to submit the decoded video sequences. The decoded video sequences shall be named like V01dec.yuv
- The participates can choose to submit a paper describing the end-to-end video compression scheme to the VCIP challenge session to or not. Please follow the VCIP paper template to prepare your manuscript if you want to submit a paper.

4. Deadlines

- August 31, registration for the competition. The authors can send the team name, team members, and the institution to lil1@ustc.edu.cn or cmjia@pku.edu.cn for registration
- September 1, release of the training and validation dataset
- September 15, submission of the challenge paper manuscript
- October 3, notification of the challenge paper acceptance
- October 10, submission of the camera-ready paper
- November 15, submission of the decoder and docker environment
- November 16, release of the test Dataset
- November 19, submission of the compressed bitstreams and decoded images/videos.
- November 30, winners and leader boards notification.
- December 4-7, challenge session at the VCIP conference. The winners will receive winner certificates provided by the VCIP organization committee. Selected teams will be invited to present at the conference.

5. Organizers

- Li Li, University of Science and Technology of China
- Chuanmin Jia, Peking University
- For any inquiries, please email us at: lil1@ustc.edu.cn; cmjia@pku.edu.cn