

No.	VC-36-N067
Resource:	IEEE 1857.11 Working Subgroup
Title:	Call for Proposals on Neural Network-Based Image Coding
Meeting Date/Place:	The 1st IEEE 1857.11 Working Subgroup Meeting / The 36th Future Video Coding Study Group Meeting Online / December 9, 2021

Summary

The IEEE 1857.11 Working Subgroup (1857.11 WSG), also known as the Future Video Coding Study Group (FVC-SG), belonging to IEEE Data Compression Standard Committee (DCSC), is executing a project to develop the IEEE Standard for Neural Network-Based Image Coding (Assigned Project Number: P1857.11). 1857.11 WSG is now calling for proposals on relevant technologies as well as standardization suggestions. The due dates for submission of a proposal are May 16, 2022 (decoder), May 19, 2022 (bitstreams and decoded images), and June 10, 2022 (technical document). This is the Call for Proposals (CfP) document as outcome of the 1st IEEE 1857.11 Working Subgroup Meeting (the 36th FVC-SG Meeting) on December 9, 2021 / online meeting.

1. Introduction

In history, neural networks had been investigated for image coding. From 1980s to 1990s, a number of studies were reported on neural network-based image compression [1, 2]. At that time, the networks were shallow and the compression efficiency was not satisfactory. Recently, neural networks, usually under the name of “deep learning,” have been reconsidered for image/video coding, which demonstrates great potential [3].

The Future Video Coding Study Group (FVC-SG), belonging to IEEE Data Compression Standard Committee (DCSC), has been studying neural network-based image/video coding for more than five years. In 2020, FVC-SG called for evidence on deep learning-based image coding technologies [4], and received several proposals. Based on the received evidence, FVC-SG developed a reference image codec, namely Neural Image Coding (NIC) [5]. It has been shown that NIC outperforms the existing standardized image coding solutions significantly in compression efficiency [6].

FVC-SG reached the consensus to develop a standard for neural network-based image coding. The standard developing activity has been approved by IEEE Standards Association and been assigned the project number P1857.11, named Standard for Neural Network-Based Image Coding. FVC-SG became the IEEE 1857.11 Working Subgroup (1857.11 WSG). 1857.11 WSG now calls for proposals on neural network-based image coding technologies as well as suggestions on neural network-based image coding standard.

The remainder of this document is organized as follows. Section 2 defines the scope of this call for proposals. Section 3 presents the detailed requirements of proposals. Section 4 is devoted to the evaluation procedure. Section 5 describes the dataset. Section 6 summarizes the timeline. Section 7 gives some other details, including contact information.

2. Scope

1857.11 WSG plans to solicit and evaluate, both objectively and subjectively, relevant solutions for neural network-based image coding. An image coding solution is considered relevant as long as it (including encoder and decoder) is built *primarily* upon neural networks and supports input images with the following attributes:

- Content: natural images (refer to the provided dataset for example);
- Resolution: up to 8K (i.e. do not exceed 7,680×4,320); internal subsampling is allowed, but decoded images shall have exactly the same resolution to the input;
- Bit-depth: 8-bit; internal bit-depth change is allowed, but decoded images should be 8-bit;
- Color space: RGB; internal color space conversion is allowed, but decoded images shall be RGB.

A solution is considered less relevant if it integrates neural network-based tools into traditional image/video coding schemes (like JPEG, JPEG2000, HEVC, VVC, etc.), e.g., a neural network-based filter to postprocess JPEG decoded images. 1857.11 WSG reserves the right to judge whether a submitted solution is relevant enough to be involved in the evaluation.

3. Requirements

Proponents shall submit a) a technical document, (b) a decoder implementation, (c) some compressed bitstreams, and (d) some decoded images, all of which are considered a complete proposal. To facilitate evaluation and to ensure fair comparison, it is required to compress the specific test data, which will be available later, without any change to the decoder. In addition, it is strongly encouraged to use the specific dataset for training and validation. More details about the dataset are provided in Section 5.

The submitted decoder, bitstreams, and decoded images may be used to verify the reproducibility of the submission. A submission failed to pass the reproducibility check will be reported by 1857.11 WSG, and the submission may not be considered as qualified. Proponents shall provide 1857.11 WSG sufficient rights to allow usage of the provided decoder for the reproducibility check. Note that every proponent who submits a complete proposal would be requested to participate in the reproducibility check, i.e., to cross-check another proposal. In addition, the submitted decoded images may be used for subjective evaluation.

More detailed requirements are specified as follows.

- Technical document. The document shall use the provided template (will be provided later). The document shall include the following information.
 - Description of the encoding and decoding algorithms. Key features of the proposal are expected to be described in detail. Internal resolution/bit-depth/color space change shall be documented if any. Especially, the structure of the used neural networks must be described in detail. Number of models and model size (including number of parameters and numerical precision of the parameters) are necessary to be reported.
 - Training data. Please report all the data that are used to train the neural networks, including the data for pretraining, for fine-tuning, and so on. It is strongly encouraged to train the neural networks with *only* the specific training data (detailed in Section 5) to facilitate comparison. Proponents are allowed to use other data, including proprietary data, for training, but this is required to be documented clearly.
 - Simulation or experimental results. Please use the provided template (will be provided later) to report ALL the necessary results obtained on the test data, including bitrate, PSNR, MS-SSIM, encoding/decoding time, etc., for each test image and each test rate, as well as a fact sheet to describe the used simulation or experimental environment. Please refer to the template for the list of the necessary results, and refer to Section 4 and Appendix A to double check how to obtain the results.
 - Standardization suggestions. Please state clearly: which part of the proposed technology is intended to be standardized; which part is normative and which part is non-normative. For the part intended to be standardized, including both normative and non-normative, please either write in or use other means to inform 1857.11 WSG chairs of the identity of each holder of any potential Essential Patent Claims of which they are personally aware.

In addition to the aforementioned mandatory information, it is highly encouraged that the technical document includes the following optional information.

- The methodology and hyperparameters used for training the networks, such as the loss function, the learning rate, the number of training epochs, etc.
- More simulation or experimental results, not restricted to the specific test data.
- Either conceptual or experimental comparison with related work.

- Ablation studies.
- Decoder implementation. A software implementation of the proposed decoding algorithm shall be provided, which allows stand-alone testing on a standard computer (equipped with GPU or not) in a reasonable amount of time. Note again that the submitted decoder is intended to be used to verify the reproducibility of the submission. Thus, the decoder shall be able to reconstruct pictures with original resolution, 8-bit, RGB, and in PNG format. To ease the reproducibility check, it is highly encouraged that the decoder is provided as a docker file and a script is provided to run the docker. The decoder can be binary executable without source code.
- Compressed bitstreams. Each test image shall be compressed to six specific bitrates, a.k.a. target rates. For each test image, the target rates are different, depending on the image content. Specifically, each test image will be compressed by BPG (see also Section 4 and Appendix B) five times with QP equal to 22, 28, 34, 40, 46, respectively, which results in five target rates; the sixth rate is lossless rate, i.e., the test image is losslessly compressed. The bitstreams shall be named like I01R03.bin, where “I01” refers to the first test image and “R03” refers to the third target rate (the lossless rate is denoted by R00). It is allowed that the actual bitrate deviates from the *lossy* target rate by no more than 10%. If the deviation is larger than 10%, the corresponding bitstream/decoded image may be submitted as well, but may be excluded in the evaluation, especially in the subjective evaluation for fairness of comparison. Bitrate deviation is not considered for the lossless rate as long as the compression is indeed lossless.
- Decoded images. For each test image and each target rate, the image decoded from the bitstream shall also be submitted. The decoded images shall be named like I01R03.png corresponding to the first test image and the third target rate.

4. Evaluation

Proposals will be evaluated from the following aspects:

- Objective evaluation of the compression efficiency. It consists of the rate-distortion performance of each proposal relative to the specified anchors, where distortion is measured by PSNR and MS-SSIM [7]. See Appendix A for more details.
- Subjective evaluation of the compression efficiency. The decoded images will be evaluated by human subjects to collect mean-opinion-score (MOS) values. For the subjective testing to be manageable, it is possible that not all decoded images will be subjectively evaluated. Specifically, a subset of test images and/or a subset of target rates will be selected by 1857.11 WSG experts, because these images/rates are believed more informative for subjective testing. In addition, significant (larger than 10%) deviation from the *lossy* target rate will result in exclusion of the corresponding decoded image. Moreover, 1857.11 WSG may choose some top performing proposals to enter the subjective testing, where the performance is decided by the objective evaluation results.
- Encoding/decoding complexity measured by the run-time of software executables.

For objective and subjective evaluations, three existing image coding solutions are used as anchors. Proponents are not required to run these anchors, as 1857.11 WSG will provide their results on the test data. Details of these anchors are provided in Appendix B.

5. Dataset

The dataset consists of three parts: training, validation, and test.

- Training set: A collection of about 1,600 pictures with high resolution are provided as the training set. The training set is publicly available at: https://structpku.github.io/LIU4K_Dataset/LIU4K_v2.html (the “Train” set therein). Normal data augmentation techniques, such as cropping, rotation, flipper over, resizing, are allowed to prepare training data. Please document the data augmentation details if any.
- Validation set: The dataset specified by NIC Common Test Conditions may be used for validation. The validation set consists of 96 images that are divided into 4 subsets, i.e. Class A (6K), Class B (4K), Class C (2K), Class D (768×512). The numbers in parentheses indicate image resolution. Each class has exactly 24 images. All these images are in RGB color space and PNG file format. The validation set may be provided to proponents upon request.
- Test set: A test set, consisting of about 20 images with high resolution, will be prepared for the evaluation. All the images are in RGB color space and PNG file format. The test set is meant to be confidential before a certain date, and will be distributed to all the registered proponents at that date. Proponents are requested to compress the test data in a specific period (no more than 72 hours).

6. Timeline

The timeline is given in the following table (*bold italic* indicates proponent's actions). Slight changes of the due dates are possible and will be announced to registered proponents.

Due Date	Item
December 21, 2021	Announcement of Call for Proposals. Release of training data and validation data.
May 15, 2022 23:59 Beijing Time	Registration of proposals (early registration is more than welcome).
May 16, 2022 23:59 Beijing Time	Submission of decoder implementation; no change of decoder is allowed after this date.
May 17, 2022 00:00 Beijing Time	Release of test data.
May 19, 2022 23:59 Beijing Time	Submission of compressed bitstreams and decoded images.
May 27, 2022	Release of objective evaluation results.
June 10, 2022 23:59 Beijing Time	Submission of technical document.
June 14, 2022	Crosscheck of reproducibility (details to be announced later).
June 17, 2022	Release of subjective evaluation results. 1857.11 WSG Meeting to discuss CfP responses.

7. Other Details

7.1 Registration, submission, and meeting

Proposals shall be registered before the specified date. Test data will be disseminated to registered proponents only. To register a proposal, please send an email in the following format to the contact detailed below.

Email format

Send to: dongeliu@ustc.edu.cn

Title: Registration of proposal

Content:

Should include the names and affiliations of all the people contributing to the proposal. Should include contact information (email and telephone number) of at least one corresponding person, who will receive the guideline to access the test data.

There will be a dedicated FTP site to submit decoder implementation, compressed bitstreams, and decoded images. FTP access information will be provided to registered proponents only. Technical document shall be submitted in the 1857.11 WSG convention.

All proponents are encouraged to participate in the 1857.11 WSG Meeting, planned on June 17, 2022, to present their proposals.

7.2 IPR conditions

This call for proposals is being made in the framework and subject to the common patent/copyright policy of IEEE and IEEE DCSC. Proponents may contact 1857.11 WSG to identify the relevant policy information.

7.3 Contact

Dong Liu, Ph.D.,

Chair of 1857.11 WSG,

Professor of the University of Science and Technology of China

Email: dongeliu@ustc.edu.cn

Mobile/WeChat: +86-181 5651 7881

References

- [1] Robert D. Dony and Simon Haykin. 1995. Neural network approaches to image compression. *Proc. IEEE* 83, 2, 288–303.
- [2] J. Jiang. 1999. Image compression with neural networks—A survey. *Signal Processing: Image Communication* 14, 9, 737–760.
- [3] Dong Liu et al. 2020. Deep learning-based video coding: A review and a case study. *ACM Computing Surveys* 53, 1, 11:1–11:35.
- [4] Future Video Coding Study Group. 2020. Call for evidence on deep learning-based image compression technologies. Tech. Rep. VC-29-N056.
- [5] https://gitlab.com/NIC_software/NIC.
- [6] NIC Software Coordinator. 2021. NIC reference software report. Tech. Rep. VC-35-M298.
- [7] Zhou Wang et al. 2003. Multiscale structural similarity for image quality assessment. 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 9-12 November 2003.

Appendix A. Distortion Measures

For objective quality assessment, four different types of distortion measures are to be used, including RGB PSNR, YUV PSNR, RGB MS-SSIM, and Y MS-SSIM. The units of all the measures are dB. These measures were defined in the NIC Common Test Conditions, and a reference implementation of calculating these measures can be found in the NIC reference software (https://gitlab.com/NIC_software/NIC/-/blob/master/Util/metrics.py).

- RGB PSNR: Mean-squared-error is calculated between the original RGB image and the reconstructed RGB image. The RGB images are flattened to vectors before calculating the mean-squared-error. Then, RGB PSNR is calculated based on the mean-squared-error.
- YUV PSNR: Images will be converted from RGB color space to YUV color space based on the BT.601 standard. For each of the three channels (Y, U, V), mean-squared-error is calculated and then PSNR is derived. Next, the three PSNR values are weighted averaged using the weights Y:U:V = 6:1:1 to obtain YUV PSNR.
- RGB MS-SSIM: For each of the three channels (R, G, B), MS-SSIM is calculated. Next, the three MS-SSIM values are directly averaged. The average value is then converted into dB scale using $-10 \times \log_{10}(1 - \text{value})$.
- Y MS-SSIM: Images will be converted from RGB to YUV based on the BT.601 standard, then the Y channel is considered to calculate MS-SSIM. The value is then converted into dB scale using the same equation as for RGB MS-SSIM.

Appendix B. Anchors

Three existing image coding solutions are used as anchors. Proponents are not required to run these anchors.

- BPG (using HEVC intra coding technologies)
 - Software to be used: BPG software version 0.9.8
 - Available at <https://bellard.org/bpg/>
 - For each test image, BPG encoder is configured with five fixed QPs: 22, 28, 34, 40, 46, to produce five target rates; moreover, BPG encoder is configured in lossless mode, to produce the lossless rate.
- VVC intra coding
 - Software to be used: VTM version 14.0
 - Available at https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM
 - RGB will be converted to YUV based on the BT.601 standard, and input to VTM with all-intra configuration; decoded YUV will be converted back to RGB.
 - For each test image and each of the five lossy target rates, VTM encoder is configured with all possible QPs, and the resulting bitrate that is most similar to the target rate is selected; moreover, RGB-YUV conversion will use a lossless transform and VTM encoder will be configured in lossless mode, to produce the lossless rate.
- NIC (MSE)
 - Software to be used: NIC version 0.5
 - Available at https://gitlab.com/NIC_software/NIC
 - For each test image and each of the five lossy target rates, NIC encoder is configured with the trained MSE-oriented variable-rate models and a set of λ values (λ is an input parameter to the variable-rate models), and the resulting bitrate that is most similar to the target rate is selected; NIC encoder does not support lossless coding, so it does not have the lossless rate.
- NIC (MS-SSIM)
 - Software to be used: NIC version 0.5
 - Available at https://gitlab.com/NIC_software/NIC
 - For each test image and each of the five lossy target rates, NIC encoder is configured with the trained MS-SSIM-oriented variable-rate models and a set of λ values (λ is an input parameter to the variable-rate models), and the resulting bitrate that is most similar to the target rate is selected; NIC encoder does not support lossless coding, so it does not have the lossless rate.