

**Joint Video Experts Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC
29/WG 11**

Document: JVET-AA0247-v4

MPEG: m60547

27th Meeting, by teleconference, 13-22 July 2022

Title: **BoG on Neural Network Video Coding (NNVC)**

Status: Input document to JVET

Purpose: Proposal

**Author(s) or
Contact(s):** [Andrew Segall](#) | asegall@amazon.com | submitter
[Elena Alshina](#) | elena.alshina@huawei.com

Source: A. Segall, E. Alshina

HEADER EXCERPT FROM SOURCE DOCUMENT

This is a report of activities from the BoG on Neural Network Video Coding (NNVC). The BoG held the following meetings during the 27th JVET meeting:

July 19 – 05:00-7:00 UTC

July 19 – 15:35-17:35 UTC

July 20 – 15:20-17:00 UTC

The BoG activity is summarized as:

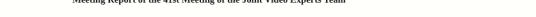
The BoG reviewed the training crosschecks, including:

Built from JVET-AA0247_BoG_v4.docx

Archive

DOC . _JVET-AA0247_BoG_v4.docx

DOC JVET-AA0247_BoG_v4.docx

  	Joint Video Experts Team (JVET) of ITU-T SG21 WP3/21 and ISO/IEC JTC 1/SC 29 27th Meeting: by teleconference, July 2022	Document: JVET-AA0247
Title:	Meeting Report of the 41st Meeting of the Joint Video Experts Team	
Status:	Report document from the chair of JVET	
Purpose:	Report	
1 Summary		
		

MEETING NOTES

[JVET-AA0247](#) BoG on Neural Network Video Coding (NNVC) [A. Segall, E. Alshina]

This is a report of activities from the BoG on Neural Network Video Coding (NNVC). The BoG held the following meetings during the 27th JVET meeting:

- July 19 – 05:00-7:00 UTC
- July 19 – 15:35-17:35 UTC
- July 20 – 15:20-17:00 UTC

The BoG activity is summarized as:

- The BoG reviewed the training crosschecks, including:
 - Summarizing key takeaways as:
 - The crosschecks took on the order of weeks for training.
 - The crosschecks took on the order of weeks for data generation. In some cases, this was further increased due to data transfer.
 - The training scripts provided by proponents did not work in all compute environments
 - The crosschecking effort would be improved by defining the naming convention for common training data, such as the number of coded frames and QPs.
 - Re-training of the model leads to a worst-case performance deviation of 0.7% in LDB between Option 1 and Option 2 training. The worst-case performance deviation for RA was 0.4%.
 - Recommendation: Consider the proposals sufficiently crosschecked
- The BoG developed recommendations for executing and reporting crosscheck information, including:
 - Developed three options for conducting a training crosscheck
 - Option 1: Generate data from scratch and train model from scratch
 - Option 2: Generate data using an existing model and train model from scratch
 - Option 3: Generate data using an existing model and initialize training using an existing model.
 - Recommendation: Allow the use of Option 1, Option 2 and/or Option 3 for future crosschecks. A crosschecker may select the most appropriate method but must report what method was used in the crosscheck report.
- The BoG developed recommendations for creating common software and methods, including:
 - Recommended that the common software have the following requirements:
 - Encoder software
 - Support for inference
 - Decoder software
 - Common data dumper with support for VTM11
 - Common data dumper with support for VTM11+reference models
 - Support for inference

- Training scripts
 - Support for Option1, Option 2 and Option 3 training
- Created a software development plan. (See table in notes below.)
- Recommendation: Develop the common software using the software development plan.
- Recommendation: Include design choices that (i) data dumper should output data in a frame based and planar format, (ii) data loader should use a JSON format for its configuration file, and (iii) software should be drift free.
- Recommendation: Franck Galpin, Yue Li, Hongtao Wang, and Liqiang Wang be the software coordinators of the common software project.
- Recommendation: Add a mandate to the NNVC AHG to develop the common software base, prepare and release anchor data for the common software base, and prepare and release verification data of the training scripts included in the common software base.
- As part of the common software discussion, the BoG also discussed the use of the common software in EE1, including:
 - Recommendation: EE should use the existing software anchor but give participants the option to use the common software when its available.
 - Recommendation: Proposals in the EE that include improvements on JVET-AA0088 and/or JVET-AA0111 be required to report results relative to these networks in addition to reporting results relative to the anchor described in the CTC.

The BoG was established with the following mandates:

- Summarize the takeaways from the training crosschecks
- Determine if common software base and/or methods can be defined
- Develop and/or refine recommendations for executing and reporting crosscheck information

Common Software Development Plan		
Task	Description	Draft Timeline
Build Environment	Include support for the SADL library in the common software and build environment	
Data dumper	Create a generic data dumper and integrate into the VTM decoder. Data dumper able to store: • Reconstructed sample value array • Prediction sample value array • Partition information array • Boundary strength information array	

	• Base QP • Slice QP • Slice type Data dumper enabled on the command line. Data dumper easily configured to store a subset of the data listed above.	
Data Loader	Create a generic data loader that is able to load data from: • DIV2K • BVI-DVC • TVD • Data dumper output Data loader easily configured to load a subset of the data provided in the data dumper output.	
Incorporate JVET-AA0088 and JVET-AA0111	Incorporate JVET-AA0088 and JVET-AA0111 into VTM encoder and decoder. Networks are based on SADL. Networks should use the same API that takes the following as input: • Reconstructed sample value array • Prediction sample value array • Partition information array • Boundary strength information array • Base QP • Slice QP • Slice type Networks should provide API that output the following: • Filtered sample value array Networks enabled through the encoder config file and command line. A flag in the SPS enables the neural network loop filter, and a second flag in the SPS selects between JVET-AA0088 and JVET-A0111. These bits indicate the use of the networks in the bit-stream for the decoder.	
Training scripts	Create training script that supports:	

	- Using the generic data loader - Training of networks in JVET-AA0088 and JVET-AA0111	
Documentation	Create documentation describing the use of the software. This should include. Description of the command line options, configuration file options, data dumper format, and API.	
Release of common software (including training scripts) rc1	Release the rc1 version of the software and training scripts for people to study	Aug. 12 2022 (3 weeks after meeting)*
Tag common software and training scripts version 1	Address potential issues such as bug fix in the rc1 versions of common software and training scripts, release version 1 of common software and training scripts	Aug. 26 2022 (5 weeks after meeting)*

**Note: The timeline is only advisory. The common software development process will prioritize software quality and completeness over meeting the draft timeline.*

Notes from JVET plenary:

It was commented that “option 3” should be considered carefully in case of crosschecks. Might initialization of a model from the proposal lead to overfitting?

It was clarified that option 3 should only be used for the training crosscheck that would verify a modification of a model that existed before, and was previously verified with option 1 or option 2.

It was clarified that also in case of option 2 the proponent will provide the crosschecker with the software that allows extraction of the data, and that the crosschecker should verify this piece of software at least with a part of the data.

A precise description of these procedures to be included in the CTC document. Review an initial version on Friday.

The recommendations of the BoG were confirmed by the JVET plenary.

It was discussed whether a separate AHG for the software should be established. It was agreed to better perform the SW development in the context of the existing AHG for the current cycle.

The possible name of the software package was discussed, but it was agreed deciding that later.

It was suggested that the AHG should also be given a mandate to discuss the structure of repository(ies). It was also reported that it is planned to develop software guidelines (input doc for next meeting ideally).

EE1 contributions: Neural network-based video coding (9)

Contributions in this area were discussed in session 8 at 1630–1730 UTC on Thursday 14 July 2022, and in session 12 at 1520–1735 UTC on Friday 15 July 2022 (chaired by JRO).

CITATION

A. Segall, E. Alshina, "BoG on Neural Network Video Coding (NNVC)", JVET-AA0247, 27th Meeting: by teleconference, July 2022.