Deep Neural Network based Video Compression for Next Generation MPEG Video Codec Standardization

This article introduces a brief status of DNNVC-related activities, current issues in the DNNVC MPEG ad-hoc group, and its standardization efforts.

Tae Meon Bae

Ofinno • ofinno.com • October 2020

1. Introduction

Video transmission is expected to dominate 79% of current global internet traffic with an annual growth rate of 31% by 2020 [1]. With this trend, the advancement of video compression technology becomes more important to alleviate radically increasing internet traffic. The Moving Picture Experts Group (MPEG) is one of the most important standardization groups working on video compression technologies.

MPEG is formed by International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) and has been leading audio and visual compression and transmission standardizations since the 1980s. The video codecs standardized by MPEG such as MPEG-2, MPEG-4, MPEG-4 Advanced Video Coding (AVC, or H.264), and High Efficiency Video Coding (HEVC or H.265) have been successfully commercialized and widely adopted. For example, 94% of over-the-top (OTT) videos were compressed by MPEG codecs in 2018 [2]. In July 2020, MPEG announced the completion of the new Versatile Video Coding (VVC or H.266) video codec. VVC generates an additional 50% compression efficiency enhancement compared with its predecessor HEVC [3].

With the finalization of the VVC standardization, MPEG experts have started discussions about a new video codec that takes advantage of Deep Neural Network (DNN) technology. MPEG established an ad-hoc group to investigate the potential of DNN-based Video Coding (DNNVC) at the 130th MPEG meeting in April 2020. Since winning the ImageNet Challenge (ILSVRC) for image classification in 2012, DNN has attracted the attention of engineers and researchers [4]. DNN illustrated outstanding performance in prediction and classification. Because prediction and classification are also particularly important in video compression, codec experts started paying attention to DNN as a promising candidate for the next generation video coding approach.

This article introduces a brief status of DNNVC-related activities, current issues in the DNNVC MPEG ad-hoc group, and its standardization efforts. To describe the latest DNNVC approaches, we referenced papers of Computer Vision and Pattern Recognition (CVPR) 2020, and a special section on learning-based image and video coding in IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) 2020. To present further insights into DNN, we also reference a presentation by Dr. Lu Yu and Dr. George Toderici at Challenge of Learned Image Compression (CLIC) 2020 [5, 6].

2. Deep Neural Network based Video Compression

From the point of view of codec architecture, two different approaches have been explored in DNNVC: Hybrid block-based coding with DNN (or Hybrid coding), and End to End learning based coding (or End to End coding). In hybrid coding approaches, DNNs replace existing encoding tools or are used as optimization methods, thereby preserving the architecture of a conventional hybrid block-based video codec. On the other hand, in End to End coding approaches, DNNs play major roles in compression, thus the architectures are highly dependent on the DNN architecture and its usage. FIG. 1 and FIG. 2 illustrate representative architectures of the hybrid coding and the End to End coding approaches.

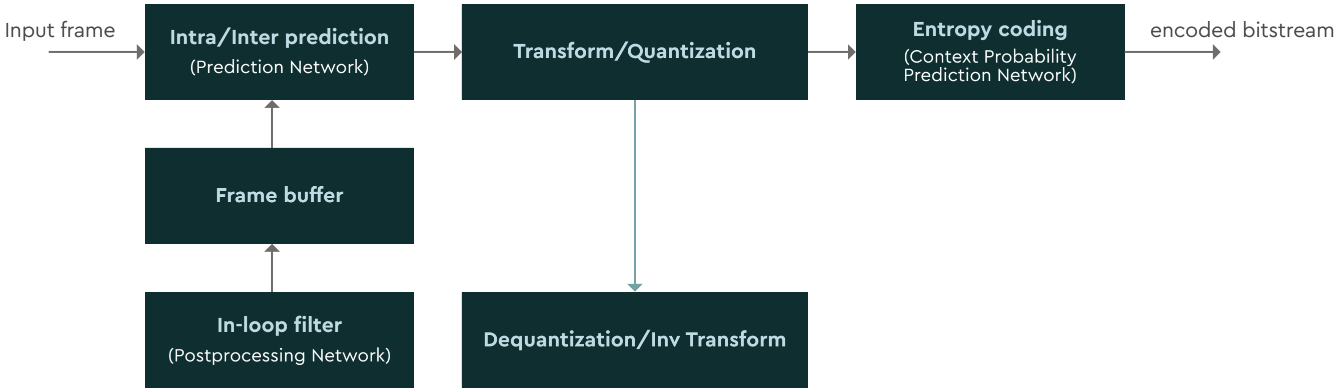


FIG. 1 Architecture of hybrid block-based coding with DNN

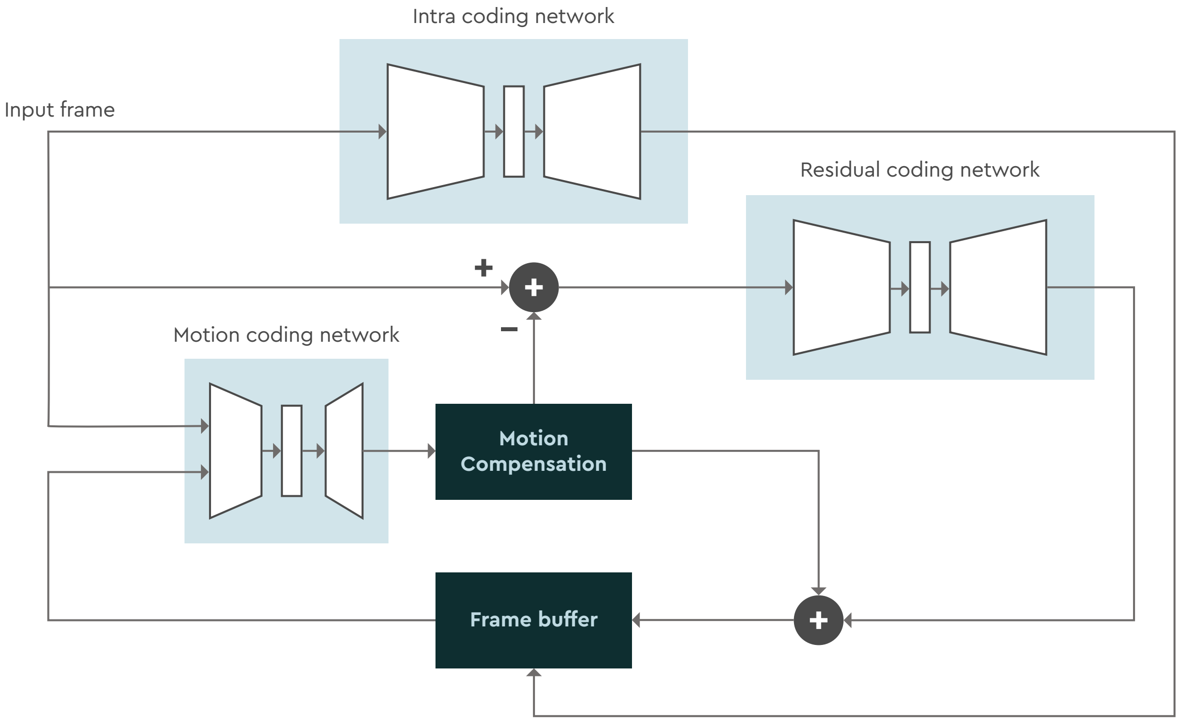


FIG. 2 An example of the End to End learning based coding architecture (Neural Video Coding)

3. Hybrid block-based coding with DNN approaches

Conventional video codecs are based on hybrid block-based coding. Main encoding tools for hybrid block-based coding include: partitioning frames into blocks, inter/intra prediction of the blocks to remove spatial/temporal redundancies, lossy compression with transform/quantization based on the human visual system, and entropy coding to remove statistical redundancy. In addition, in-loop filtering such as deblocking and Sample Adaptive Offset (SAO) filters are used to reduce compression artifacts. In hybrid coding approaches, DNNs are used for estimations such as: intra/inter prediction and compression artifact removal. In entropy coding, DNN is used to predict probability (or possibility) of contexts for Context Adaptive Binary Arithmetic Coding (CABAC). FIG. 3 shows performance enhancements with DNN based encoding tools compared with HEVC [5]. DNN based Intra/Inter prediction and entropy coding achieved an additional 3~5% bitrate saving. Among the DNN based encoding tools, DNN based in-loop filtering showed the highest coding enhancements up to 8.6%, and the deeper network showed a higher enhancement [5]. One of the hybrid coding approaches named Deep Learning-based Video Coding (DLVC) was proposed in the 122nd MPEG meeting in 2018. DLVC added two deep Convolutional Neural Network (CNN) based coding tools on top of HEVC: a convolutional neural network-based loop filter (CNNLF), and a convolutional neural network-based block adaptive resolution coding (CNN-BARC). DLVC showed a 33~39.6% coding improvement compared with HEVC [7]. A second hybrid coding approach with DNN was proposed by Lu Yu in an MPEG 131 meeting [8]. Its aim is to improve the accuracy of bi-prediction using CNN. This second hybrid coding approach showed a 2.92~5.06% bit saving compared with VVC, and it was also selected as one of the winners for p-frame compression in the CLIC 2020 hosted by CVPR 2020 [9].

4. End to End learning based coding approaches

Compared with the hybrid coding approaches, there is no concrete consensus about a common End to End learning based coding approach until now. But most End to End learning based coding approaches also try to remove spatial/temporal redundancy and utilize lossy compression with quantization and entropy coding like the hybrid coding approaches. An End to End approach named NVC (Neural Video Coding) was proposed in an MPEG 131 meeting in 2020 [10]. As seen in FIG 2, NVC comprises intra, motion, and residual coding networks. In addition, quantization and arithmetic coding is adopted with a context prediction network (named hyper coder). FIG. 4 shows that the NVC coding performance has a better coding gain than HEVC [10]. A comparison of NVC with VVC was not presented. However, it is possible to compare the performance of the End to End coding approaches with that of VVC from the

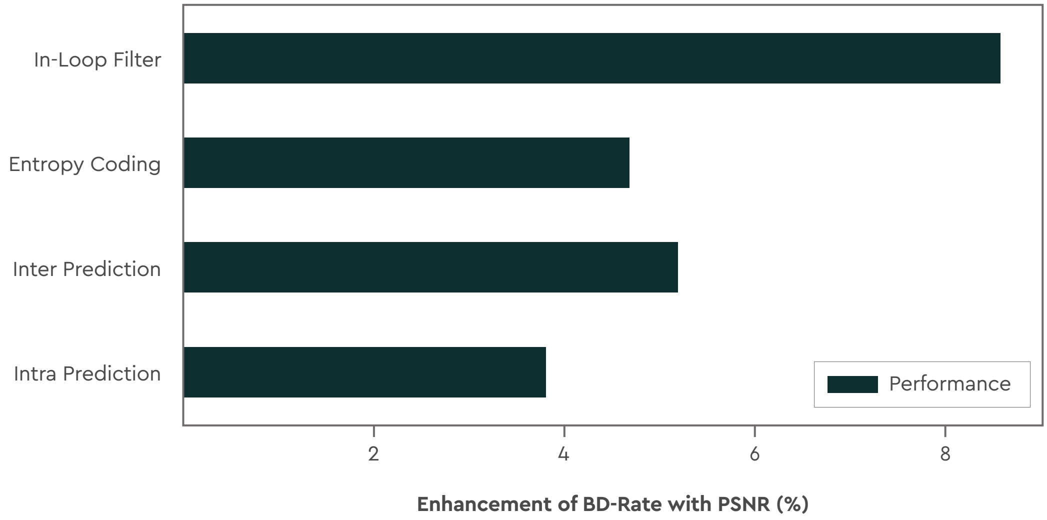


FIG. 3 Performances of DNN encoding tools

results of CLIC 2020. In CLIC 2020, 2 of 3 winners of p-frame compression are End to End coding approaches. They showed slightly better coding efficiencies than VVC [11].

5. Activities and prospects for DNNVC ad-hoc group in MPEG

The DNNVC ad-hoc group started in April 2020. In July 2020, the main discussions in the ad-hoc group were about test conditions for evaluation of different proposals. In addition to the conventional tests, training, and test environments of neural networks were discussed. For example, in the MPEG standardization, experimental results of proposals are crosschecked by other experts. But the randomness in DNN training, which is also being discussed, makes this process difficult.

Another active discussion topic was about visual quality metrics. The Peak Signal-to-Noise Ratio (PSNR) and Multiscale Structural Similarity (MS-SSIM) are widely used objective visual quality metrics. But they show discrepancies with human perceived visual quality. These discrepancies make it difficult to compare coding performance between different codecs. In addition, a precise perceptual visual quality metric could help enhancing coding efficiency. Especially in the case of DNNVC, a precise perceptual visual quality metric could be important because it is widely used as a loss function of the neural network. Learning based visual quality metrics such as the Learned Perceptual Image Patch Similarity (LPIPS) metric was mentioned in the meeting as a candidate for a visual quality metric [12].

About the architecture point of view, both of Hybrid block-based coding with DNN (or Hybrid coding), and End to End learning based coding (or End to End coding) will be considered in the ad-hoc group.

6. Conclusion

Recent research showed that deep neural network-based video compression is very promising. VVC has just ended and so it will take time to finish next generation video codec standardization. DNNVC needs to be enhanced to outperform VVC with a reasonable increase in computational complexity. For each MPEG codec generation, coding efficiency has been enhanced by more than 50%. DNNVC also has to prove a similar achievement compared with VVC.

As mentioned, several DNN based encoding tools were proposed but not accepted during VVC standardization due to their high computational

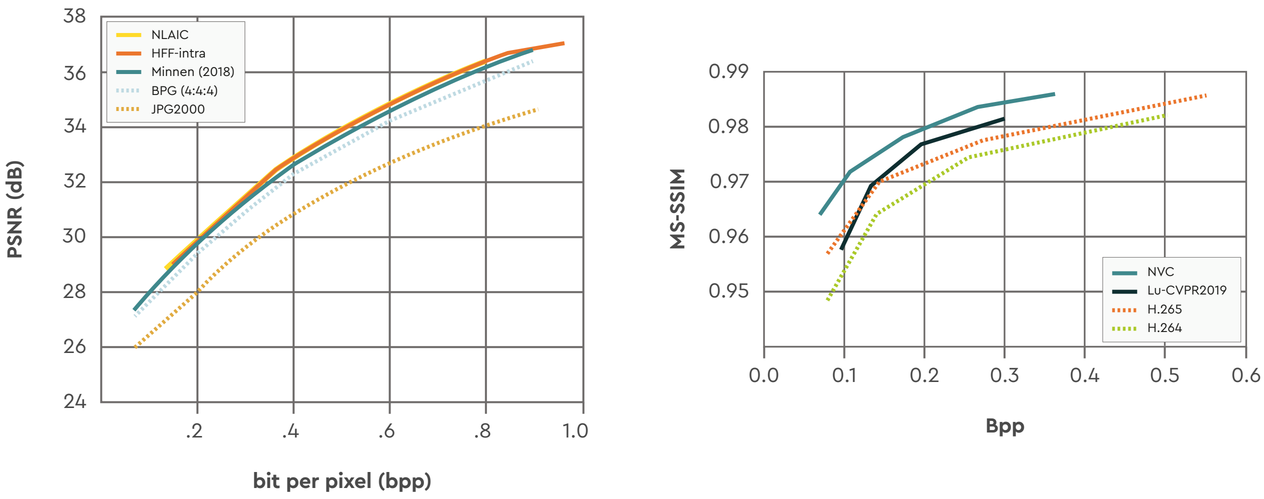


FIG. 4 NVC coding performance

costs. Proponents were also concerned about computational complexity, so they did not design their networks deeply enough. This resulted in marginal coding gain compared with non-DNN based approaches at the time. The computational complexity of DNN is still an important issue. The winning codecs of CLIC 2020 are also computationally expensive. Even considering H/W advancements for DNN in the near future, computational complexity will be an important problem that should be overcome.

To succeed in DNNVC standardization, a common codec architecture for DNNVC needs to be considered. Even if DNNVC achieves great performance, standardization will be difficult if DNNVC cannot include various contributions in one architecture. These topics are currently being discussed in DNNVC ad-hoc group.

7. Acronyms

CLIC	Challenge of Learned Image Compression
CVPR	Computer Vision and Pattern Recognition
CABAC	Context Adaptive Binary Arithmetic Coding
CNN	Convolutional Neural Network
CNN-BARC	Convolutional neural network-based block adaptive resolution coding
CNNLF	Convultional neural network-based loop filter
DNNVC	DNN-based Video Coding
DLVC	Deep Learning-based Video Coding
DNN	Deep Neural Network
HEVC	High Efficiency Video Coding
ILSVRC	ImageNet Challenge
ISO/IEC	International Organization for Standardization/International Electrotechnical Commission
LPIPS	Learned Perceptual Image Patch Similarity
AVC	MPEG-4 Advanced Video Coding
MPEG	Moving Picture Experts Group
MS-SSIM	Multiscale Structural Similarity
NVC	Neural Video Coding
PSNR	Peak Signal-to-Noise Ratio
SAO	Sample Adaptive Offset
TCSVT	Transactions on Circuits and Systems for Video Technology
VVC	Versatile Video Coding
OTT	Over-the-Top

References

1. “CISCO Global—Forecast Highlights,” 2020, pp. 2.

2. “Reports Show HEVC Usage on the Rise,” Streaming media, https://www.streamingmedia.com/Articles/News/Online-Video-News/Reports-Show-HEVC-Usage-on-the-Rise-135776.aspx?utm_source=related_articles&utm_medium=gutenberg&utm_campaign=editors_selection

3. “Press Release of 131st WG11 Meeting,” w19387, https://mpeg-standards.com /meetings/ mpeg -131/

4. “Result of Large Scale Visual Recognition Challenge 2012 (ILSVRC2012),” http://image-net.org/challenges/LSVRC/2012/results.html#t1

5. https://www.youtube.com/watch?v=rfODWzvDWic, Lu Yu (yul@zju.edu.cn)

6. https://www.youtube.com/watch?v=iXzgFrRWNEg&feature=youtu.be, George Toderici

7. “Description of SDR video coding technology,” JVET-J0032, 10th MPEG, San Diego, Apr., 2018

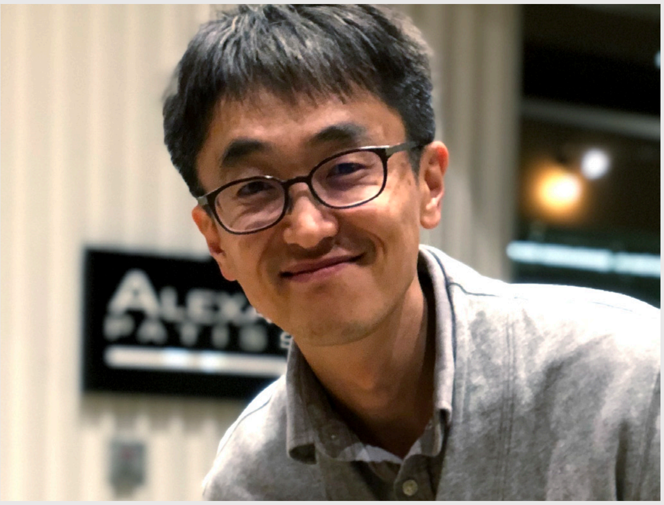
8. “Bi-prediction with CNN Utilizing Spatial and Temporal Information for High-Efficiency Video Coding,” m54375, 131th MPEG meeting, online, June 2020

9. CLIC web page, http://compression.cc

10. “End-to-End Neural Video Coding – From Pixel Prediction to Feature Sensing,” m54341, 131th MPEG meeting, online, June 2020

11. “Analysis results of Workshop and Challenge on Learned Image Compression in CVPR2020,” m54556, 131th MPEG meeting, online, June 2020

12. “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” Richard Zhang et al, CVPR 2018



About the Author:

Tae Meon Bae is a research engineer offering over 20 years of video technology expertise. He has a significant level of experience working on machine learning for video understanding, video compression and streaming in academic and industrial domains. He has an impressive 112 international patent applications and 36 published international papers. From 2011 to 2019, he has led R&D and commercialization of Cloud Streaming (CS): he is the world's first inventor of the CS based Set Top Box App (STB) Platform and succeeded world's first commercialization of CS based STB App Service in SK Telecom/SK Planet in South Korea. His current research interests include Deep Neural Network based video compression and codec optimization.

About Ofinno:

Ofinno, LLC, is a research and development lab based in Northern Virginia, that specializes in inventing and patenting future technologies. Ofinno's researchers create technologies that address some of the most important issues faced by wireless device users and the carriers that serve them. Ofinno's inventions have an impressive utilization rate. Ofinno's research involves technologies such as 5G Radio and Core networks, IoT, V2X, and ultra-reliable low latency communications. Our innovators not only create the technologies, they oversee the entire process from the design to the time the technology is sold. For more information about Ofinno, please visit www.ofinno.com.