딥러닝을 활용한 비디오 코딩 기술 표준화 동향

김연희
ETRI 선임연구원

GISC2020
Global ICT Standards Conference

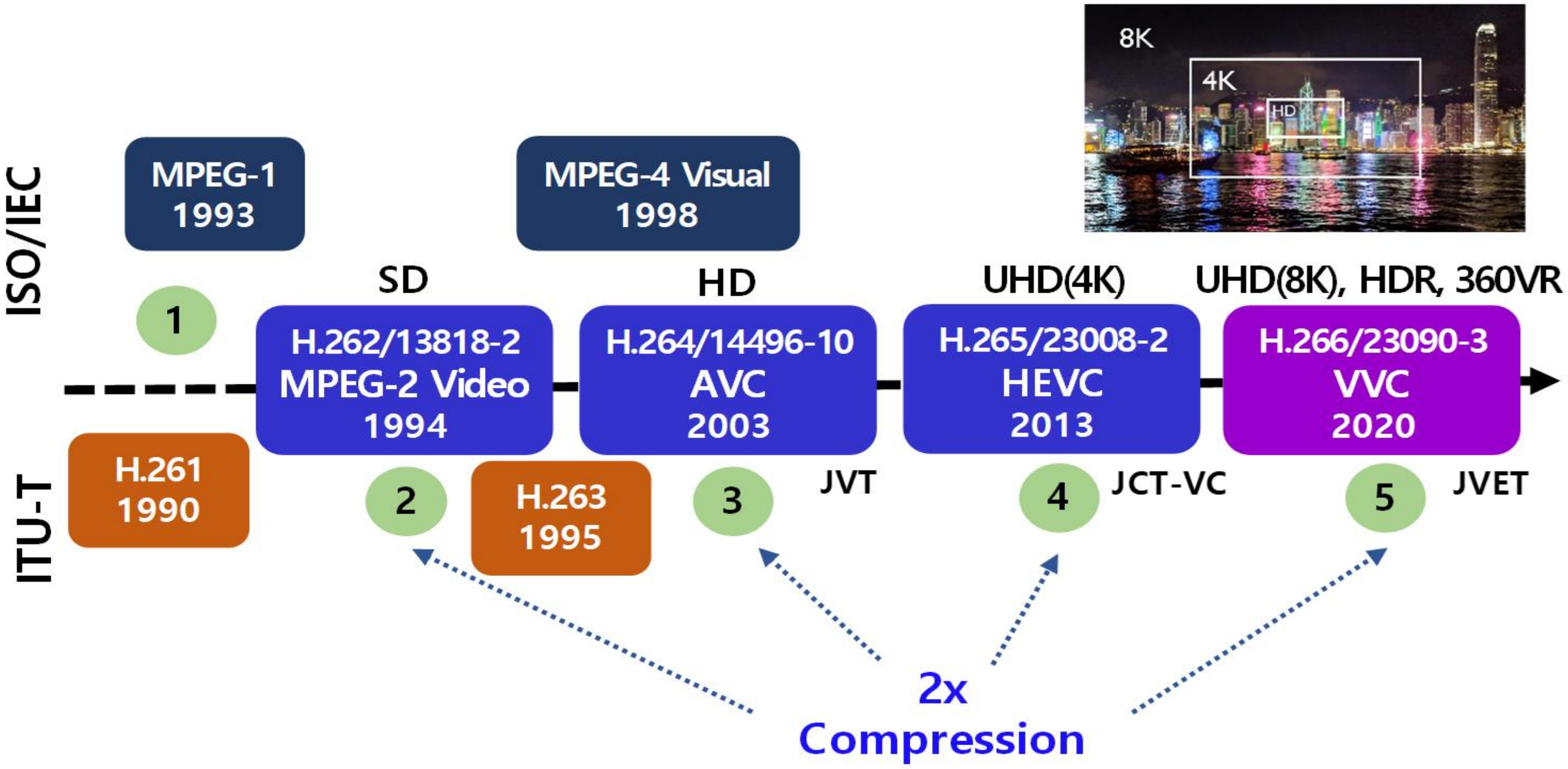
뉴 노멀 시대
선도를 위한
ICT 표준의
역할

목차

- Introduction
- NN-base video coding approaches
- MPEG DNNVC
- JVET NNVC
- Conclusion

Introduction

- Video coding standards, double compression ratio every 10 years (or so)



ISO/IEC

MPEG-1
1993

MPEG-4 Visual
1998

ITU-T

H.261
1990

H.263
1995

SD

H.262/13818-2
MPEG-2 Video
1994

HD

H.264/14496-10
AVC
2003

JVT

UHD(4K)

H.265/23008-2
HEVC
2013

JCT-VC

UHD(8K), HDR, 360VR

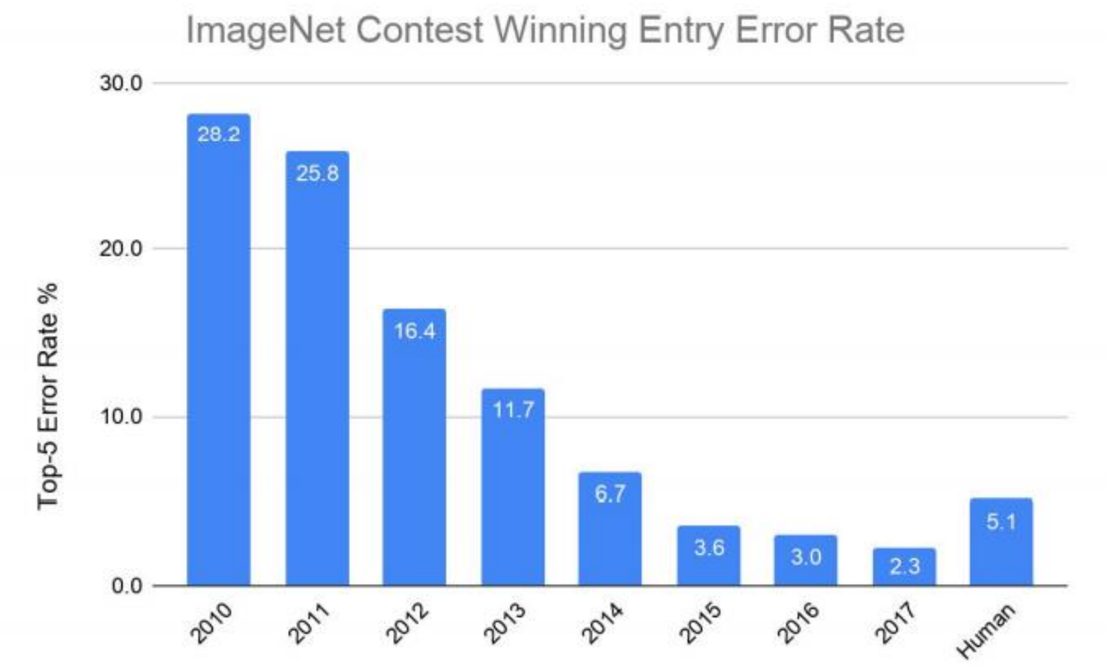
H.266/23090-3
VVC
2020

JVET

2x
Compression

Introduction

- The deep learning revolution
- What about video coding?



< ImageNet 영상 인식 오류율[1] >

[1] **J. Dean**, "The Deep Learning Revolution and Its Implications for Computer Architecture and Chip Design,", 2020

GISC2020
Global ICT Standards Conference

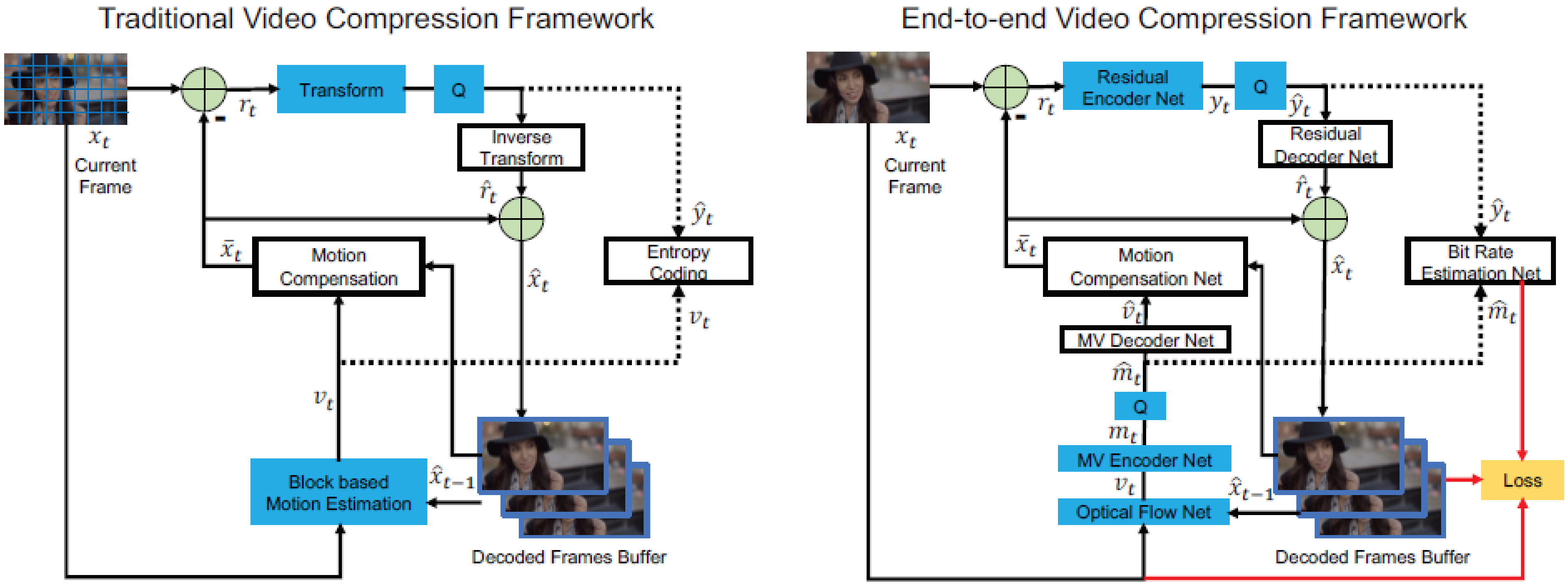
뉴 노멀 시대
선도를 위한
ICT 표준의
역할

목차

- Introduction
- NN-base video coding approaches
- MPEG DNNVC
- JVET NNVC
- Conclusion

NN-based video coding approach

- End-to-end approach



Traditional Video Compression Framework

End-to-end Video Compression Framework

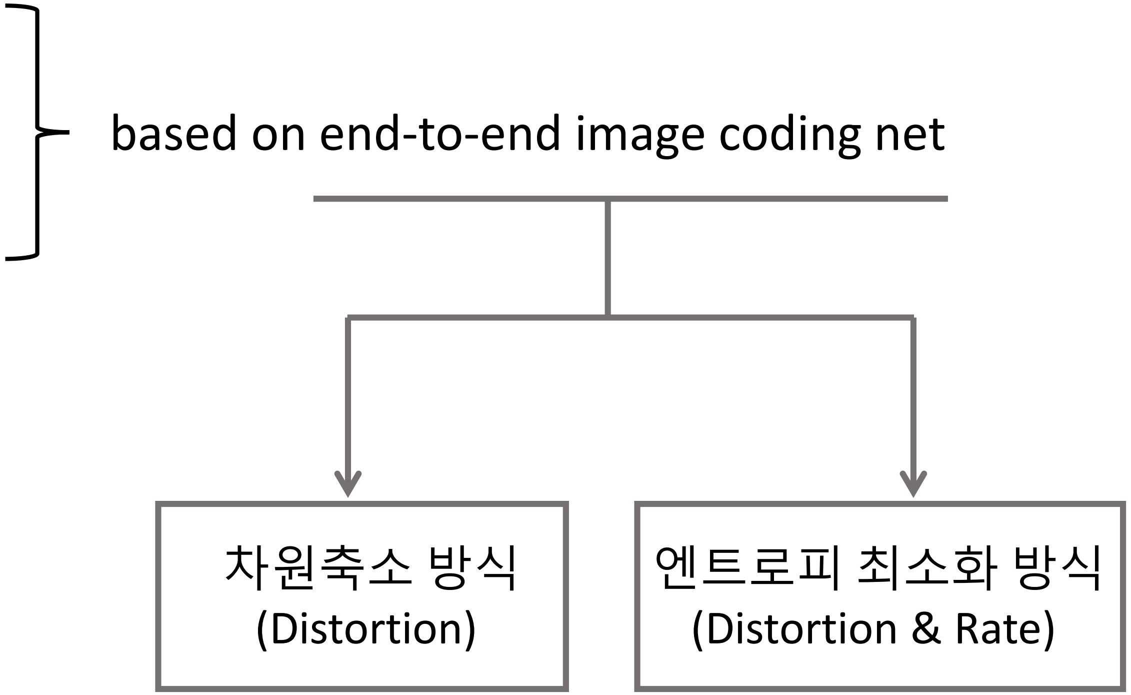
< 기존 비디오 코딩 구조와 End-to-end 압축 구조[2] >

[2] **G. Lu,** et al., "DVC: An end-to-end deep video compression framework," 2019

NN-based video coding approach

- End-to-end approach

  - Optical flow net
  - MV encoder net
  - Residual encoder net
  - Bit-rate estimation net



based on end-to-end image coding net

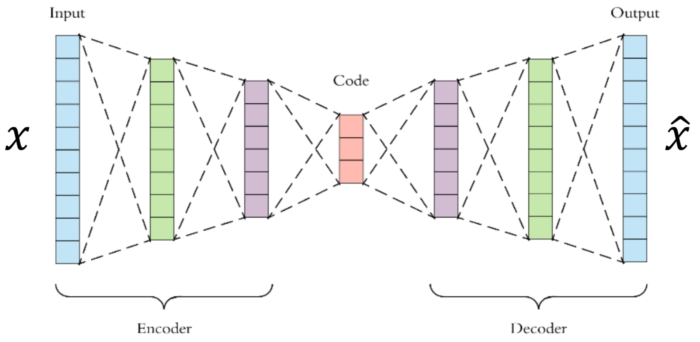
차원축소 방식
(Distortion)

엔트로피 최소화 방식
(Distortion & Rate)

NN-based video coding approach

- End-to-end approach
  - Based on end-to-end image coding net

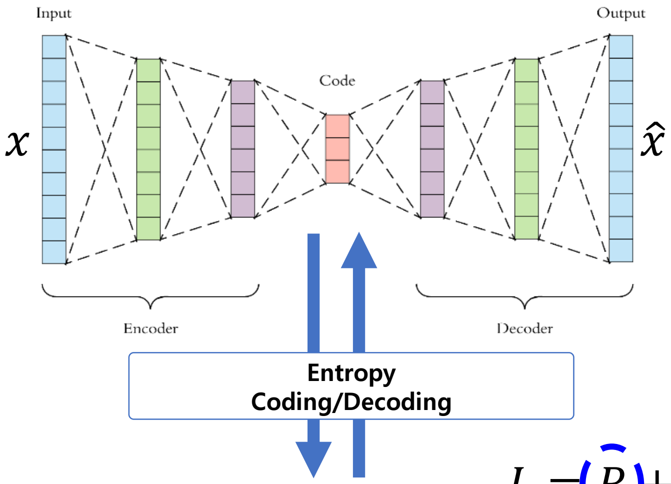
차원축소 방식
(Distortion)



$$L = \underline{\mathrm{D}(x,\hat{x})}$$

MSE, MS-SSIM, etc.

엔트로피 최소화 방식
(Distortion & Rate)



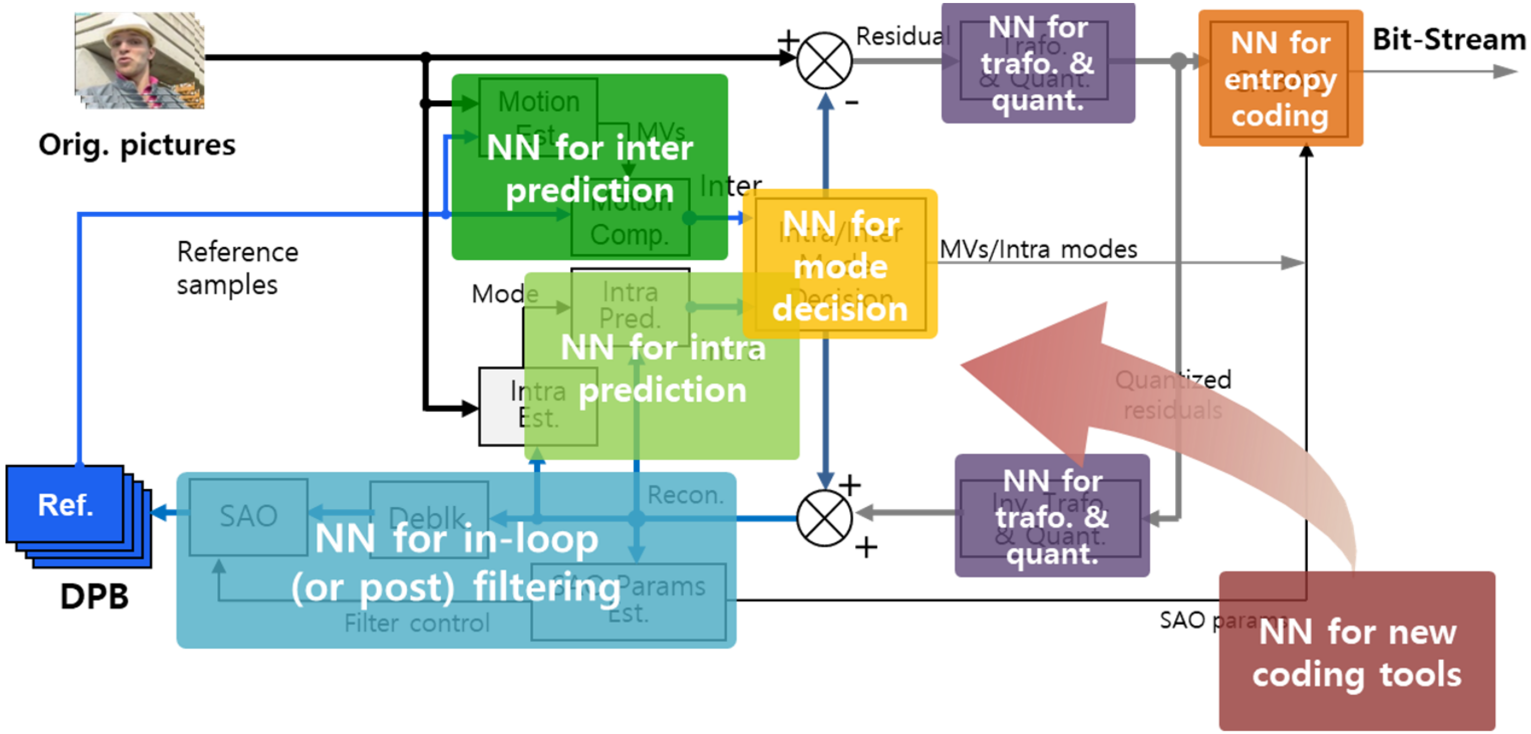
01101110001…
bitstream

$$L = \boxed{R} + \lambda\,\mathrm{D}(x,\hat{x})$$

엔트로피로 대체하여 계산

NN-based video coding approach

- Tool-by-tool approach



< 기존 비디오 코딩 모듈 별 NN-based approach>

GISC2020
Global ICT Standards Conference

뉴 노멀 시대
선도를 위한
ICT 표준의
역할

NN-based video coding approach

- Tool-by-tool approach
  - Intra Prediction
  - Inter Prediction
  - In-loop Filtering

NN-based video coding approach

- Tool-by-tool approach
  - Intra Prediction

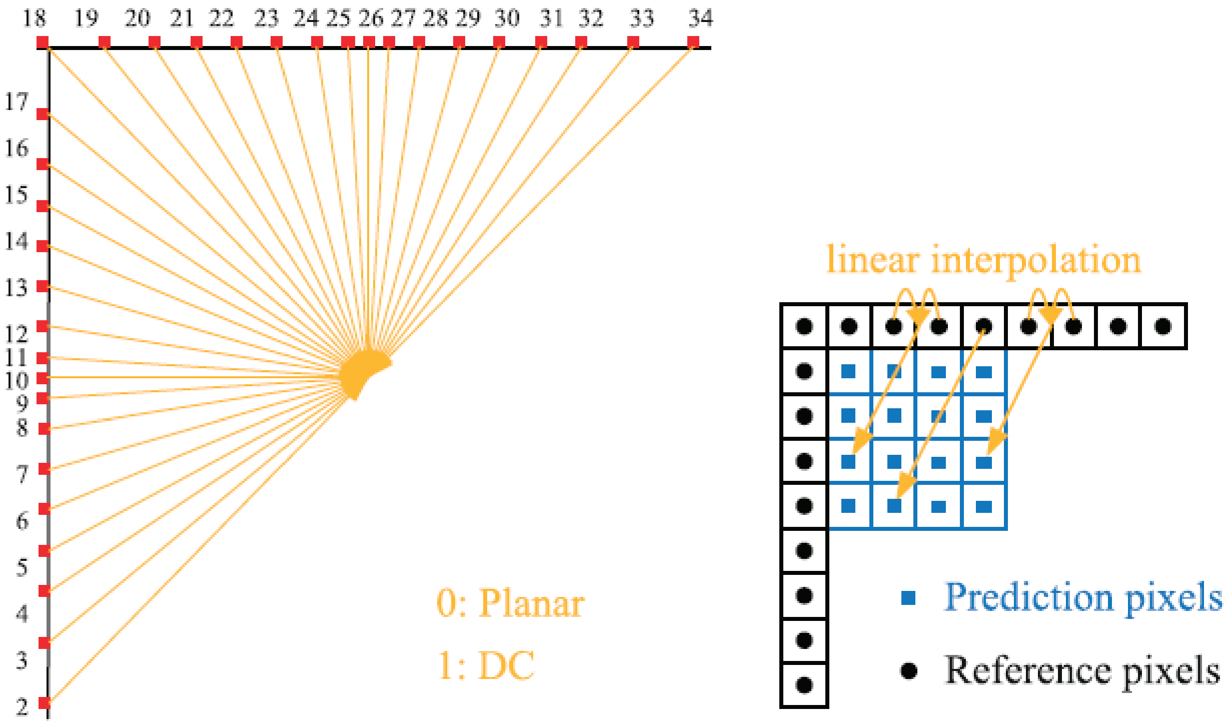
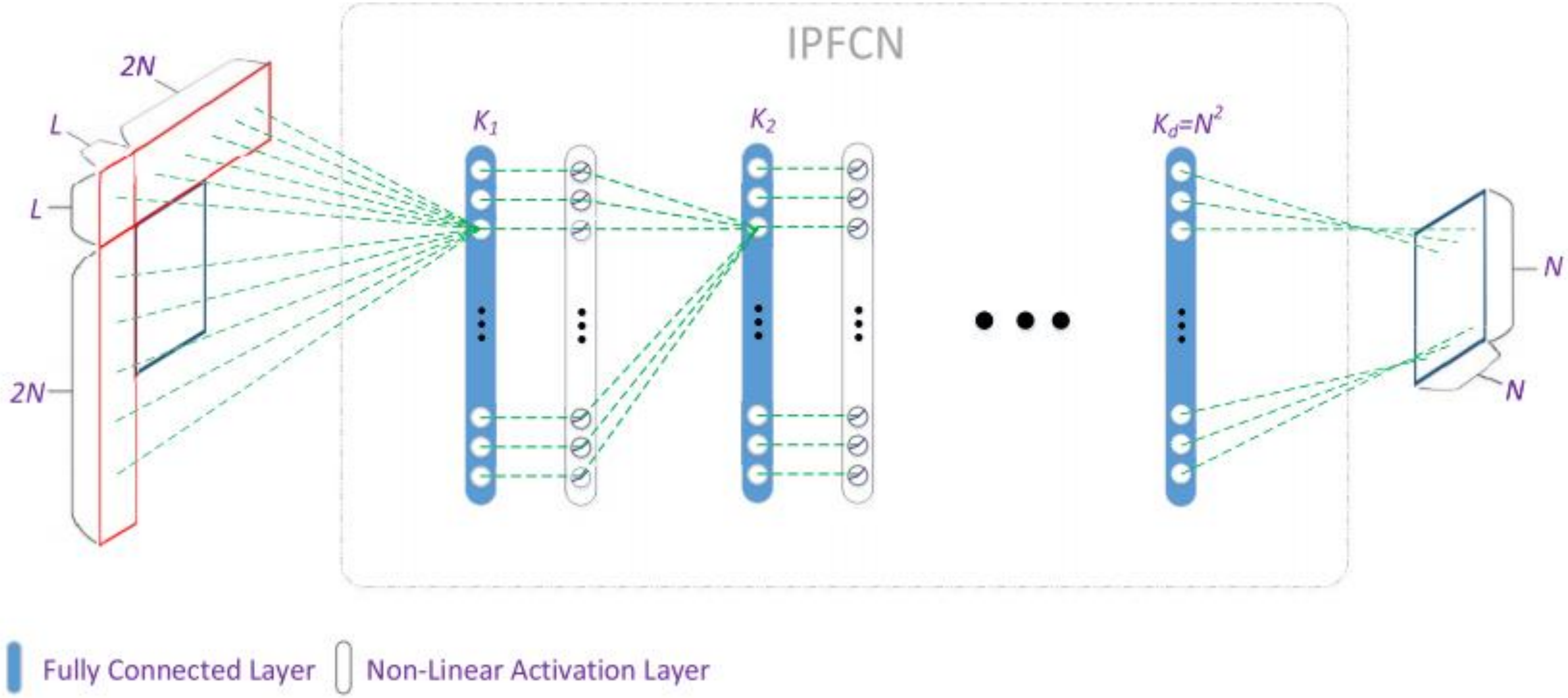


Fig. 1. HEVC intra prediction. (Left) 35 intra prediction modes. (Right) An illustration of HEVC intra angular prediction.

NN-based video coding approach

- Tool-by-tool approach
  - Intra Prediction



[3] **J. Li,** et al., "Fully Connected Network-Based Intra Prediction for Image Coding," 2018

NN-based video coding approach

- Tool-by-tool approach
  - Inter Prediction

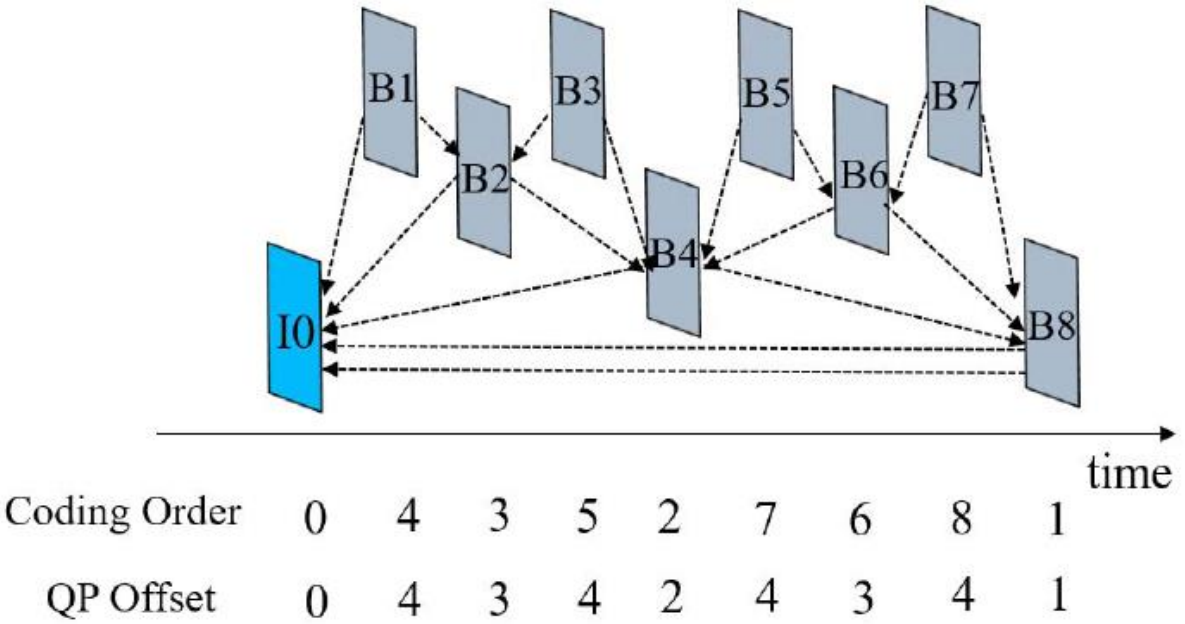
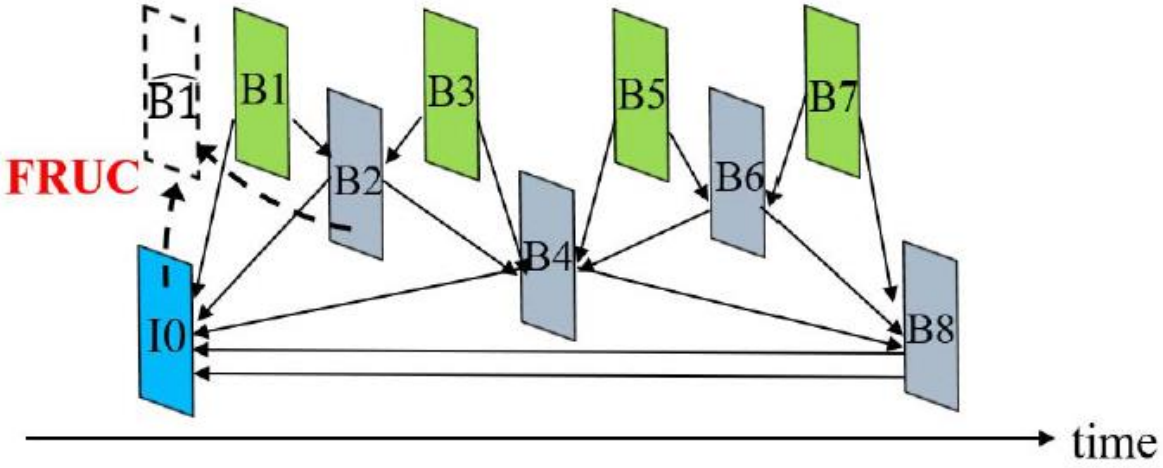


Fig. 1. Hierarchical B coding structure in HEVC

NN-based video coding approach

- Tool-by-tool approach
  - Inter Prediction



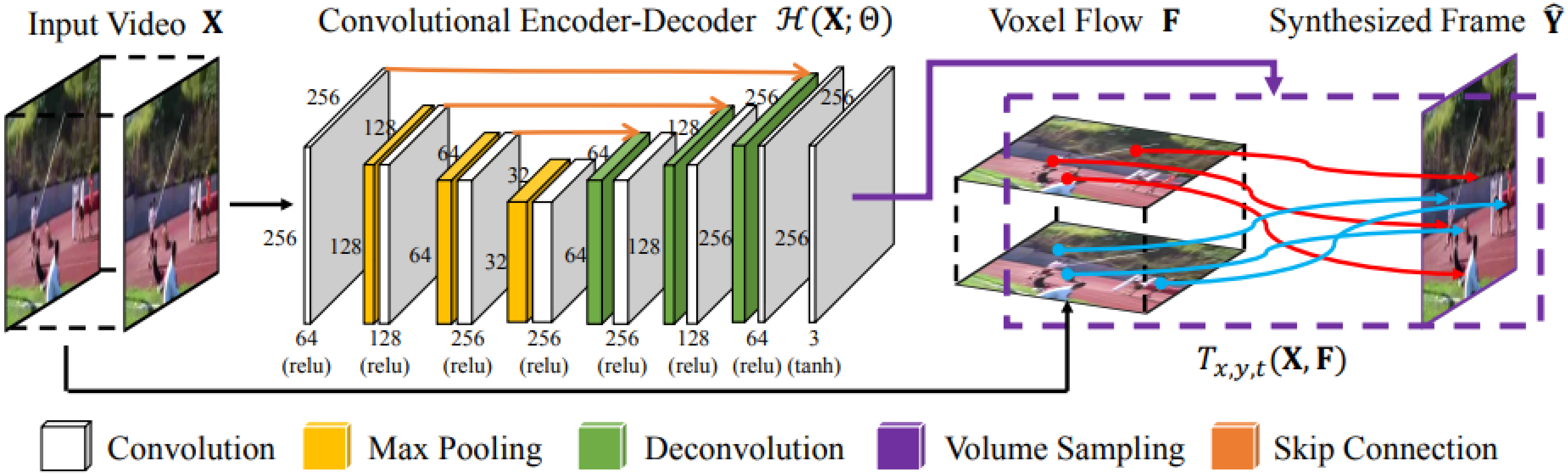
Coding Order	0	4	3	5	2	7	6	8	1
QP Offset	0	4	3	4	2	4	3	4	1

Fig. 4. Illustration of VRF generation with hierarchical B coding structure.

[4] **L. Zhao,** et al., “Enhanced Motion-compensated Video Coding with Deep Virtual Reference Frame Generation,” 2019.

NN-based video coding approach

- Tool-by-tool approach
  - Inter Prediction
    - Deep FRUC is based on [5]



[5] **Z. Liu,** et al., “Video frame synthesis using deep Voxel flow,” 2017.

NN-based video coding approach

- Tool-by-tool approach
  - In-loop Filtering

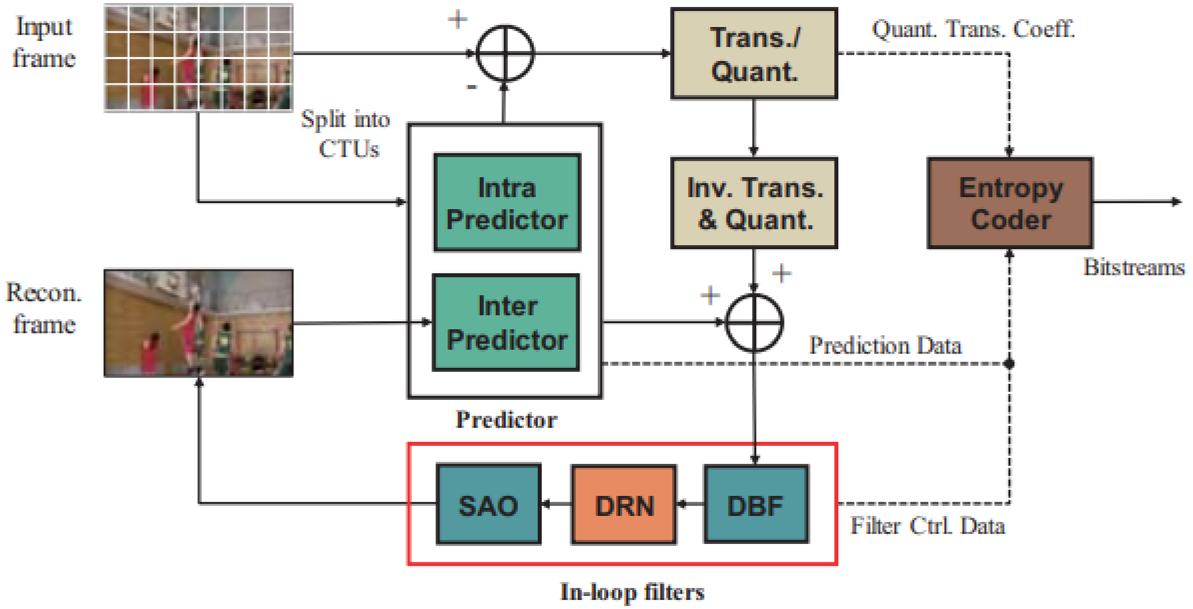
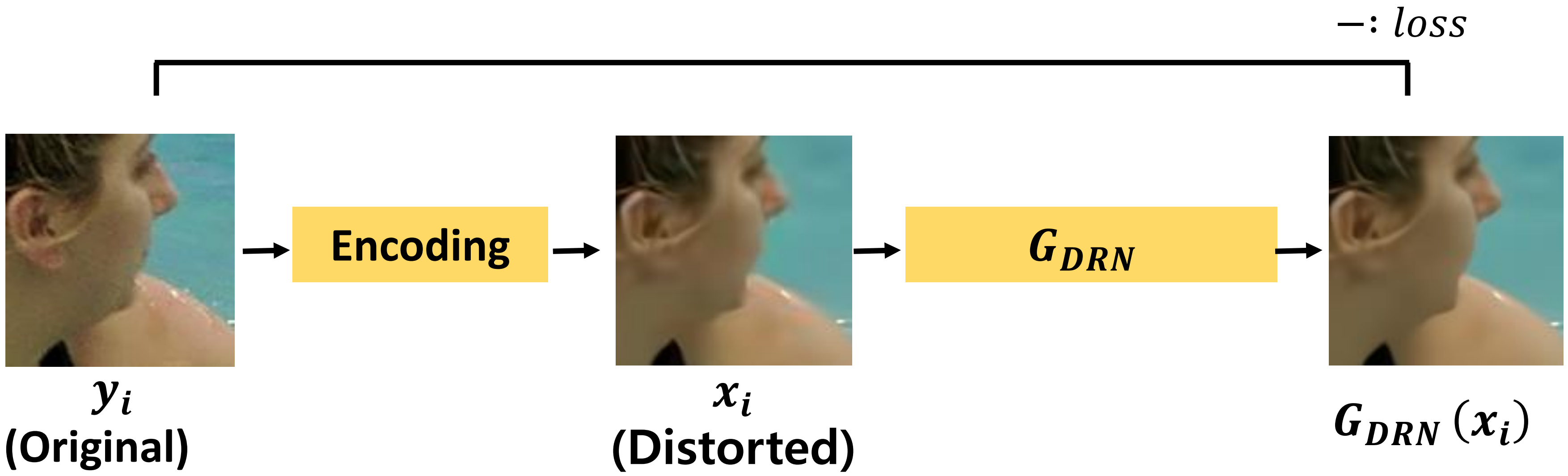


Fig. 3: Integration of DRN based in-loop filter into HEVC.

[6] **Y. Wang,** et al., "Dense Residual Convolutional Neural Network based In-Loop Filter for HEVC," 2018.

NN-based video coding approach

- Tool-by-tool approach
  - In-loop Filtering



$$L = \mathbb{E}[\|G_{DRN}(x_i) - y_i\|]$$

목차

- Introduction
- NN-base video coding approaches
- MPEG DNNVC
- JVET NNVC
- Conclusion

DNNVC: 개요

- Established at the 130th meeting
  - Up-to-date report of the current progress and status for deep network based approaches in the field of image and video coding (m53700)
  - Two categories: neural networks in hybrid coding framework and end-to-end optimized coding
- Chairmen
  - Yan Ye (chair), Elena Alshina, Jianle Chen, Shan Liu, Jonathan Pfaff, Shanshe Wang (co-chairs)
- Email reflector
  - mpeg-dnnvc@lists.aau.at via https://lists.aau.at/mailman/listinfo/mpeg-dnnvc

DNNVC: 개요

- Mandates
  1. Evaluate and quantify performance improvement potential of DNN based video coding technologies (including hybrid video coding system with DNN modules and end-to-end DNN coding systems) compared to existing MPEG standards such as HEVC and VVC, considering various quality metrics;
  2. Study quality metrics for DNN based video coding;
  3. Solicit input contributions on DNN based video coding technologies;
  4. Analyze the encoding and decoding complexity of NN based video coding technologies by considering software and hardware implementations, including impact on power consumption;
  5. Investigate technical aspects specific to NN-based video coding, such as design network representation, operation, tensor, on-the-fly network adaption (e.g. updating during encoding) etc

131차 MPEG DNNVC AHG

- 3 sessions of AHG meeting
  - Participants: 167 ~ 241
- 10 input document
  - Performance evaluation: 2
  - Quality metrics: 1
  - Coding technologies: 4
  - Others: test conditions and platforms: 3

GISC2020
Global ICT Standards Conference

뉴 노멀 시대
선도를 위한
ICT 표준의
역할

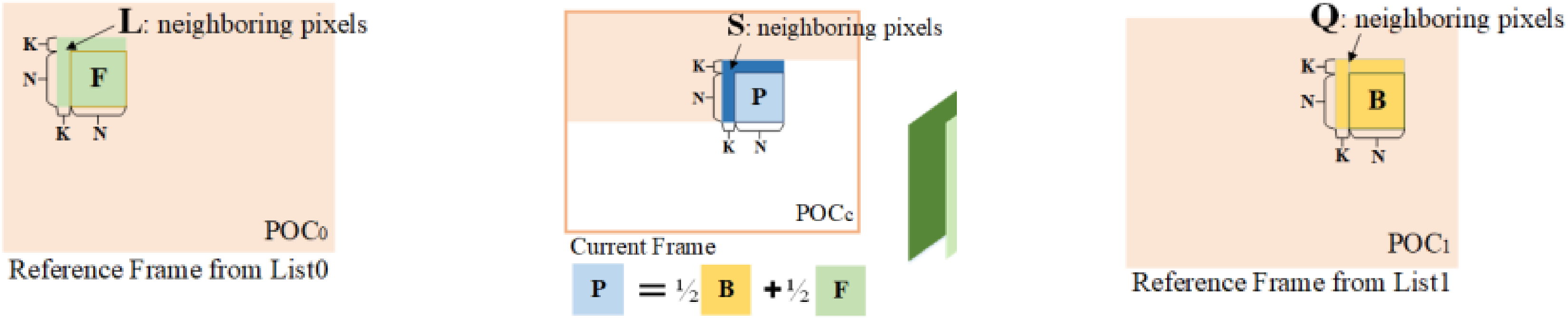
131차 MPEG DNNVC AHG

- 10 Input Contributions

Doc No.	Title	Authors
m54341	[DNNVC] End-to-End Neural Video Coding - From Pixel Prediction to Feature Sensing	H. Liu, T. Chen, M. Lu, Z. Ma (Nanjing University), Y. Wang (New York University), L. Yang (Gyrfalcon Technology), Z. Xie, F. Wang, D. Wang (OPPO), L. Wang (Hikvision Research Institute)
m54375	[DNNVC] Bi-prediction with CNN Utilizing Spatial and Temporal Information	Lu Yu, Yule Sun (Zhejiang University), Jue Mao, Haitao Yang (Huawei)
m54381	[DNNVC] Learned Visual Quality Assessment for Image and Video Compression	Jiaqi Zhang, Lu Yu (Zhejiang University)
m54452	[DNNVC] proposed common test conditions for deep neural network based video coding	Y. Ye, Z. Wang, R.-L. Liao (Alibaba), S. S. Wang, S. W. Ma (PKU)
m54453	[DNNVC] Study of Deep Learning Frameworks	Z. Wang, R.-L. Liao, Y. Ye (Alibaba), S. S. Wang, S. W. Ma (PKU)
m54467	[DNNVC] A PyTorch library and evaluation platform for end-to-end compression research	J. Bégaint, F. Racapé, S. Feltman, A. Pushparaja (Interdigital)
m54556	[DNNVC] Analysis results of Workshop and Challenge on Learned Image Compression in CVPR2020	E.Alshina, N. Giuliani, A.Karabutov, H. Chen, X. Ma, H. Yang (Huawei)
m54586	[DNNVC] Bit allocation analysis of learning-based image compression	N. Yan, D. Liu, H. Li, F. Wu (USTC), N. Song, H. Yang (Huawei)
m54541	[DNNVC] Block based learned image compression	Z.H. Zhao, C.M. Jia, S.S. Wang, S.W. Ma (PKU), Z. Wang, Y. Ye (Alibaba)
m54739	[DNNVC] Substitutional Neural Image Compression	W. Wang, W. Jiang, X. Wang, S. Liu (Tencent)

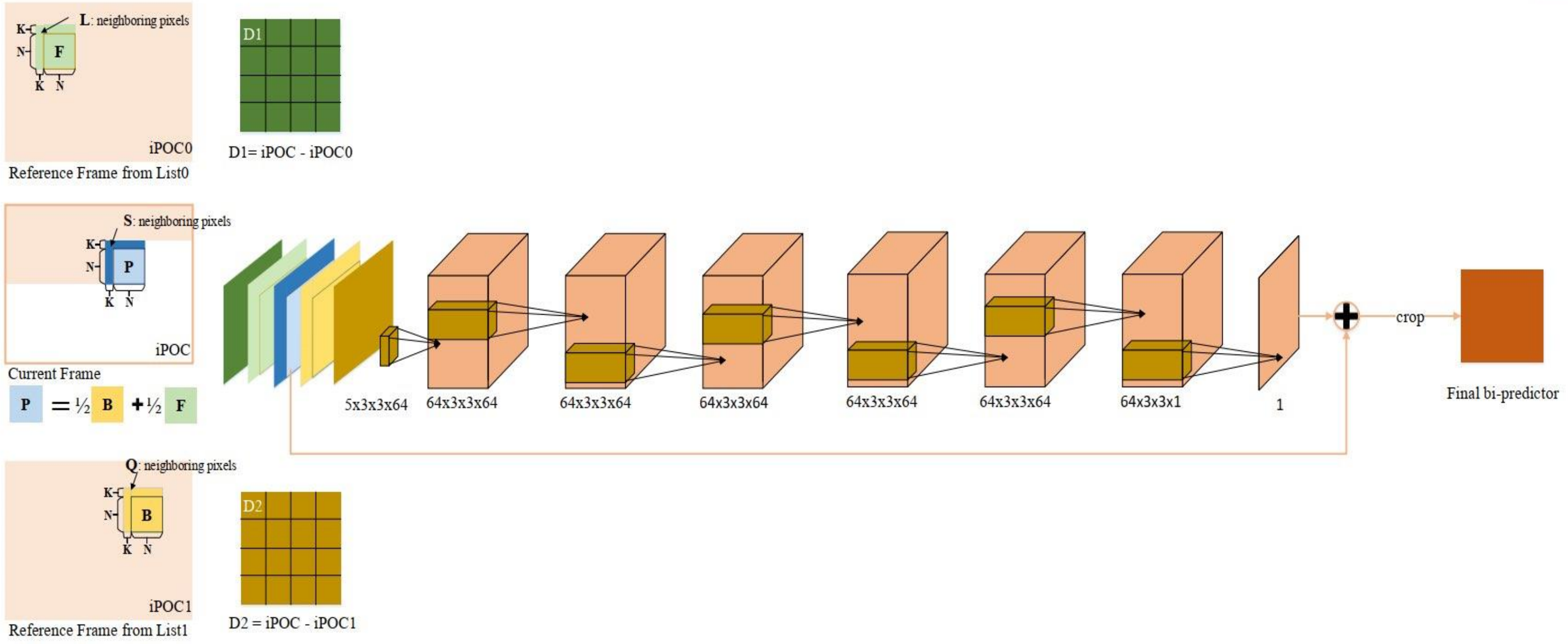
131차 MPEG DNNVC AHG: Input document 1 (1/3)

- [m54375] Bi-prediction with CNN Utilizing Spatial and Temporal Information of HEVC
  - To improve the accuracy of bi-prediction
  - Input:
    - two reference blocks
    - spatial neighboring pixels of both the current block and two reference blocks
    - temporal distances between the current block and the reference blocks
  - Output:
    - estimate the temporal variation of pixel values
    - estimate the similarities between the reference blocks and the current block



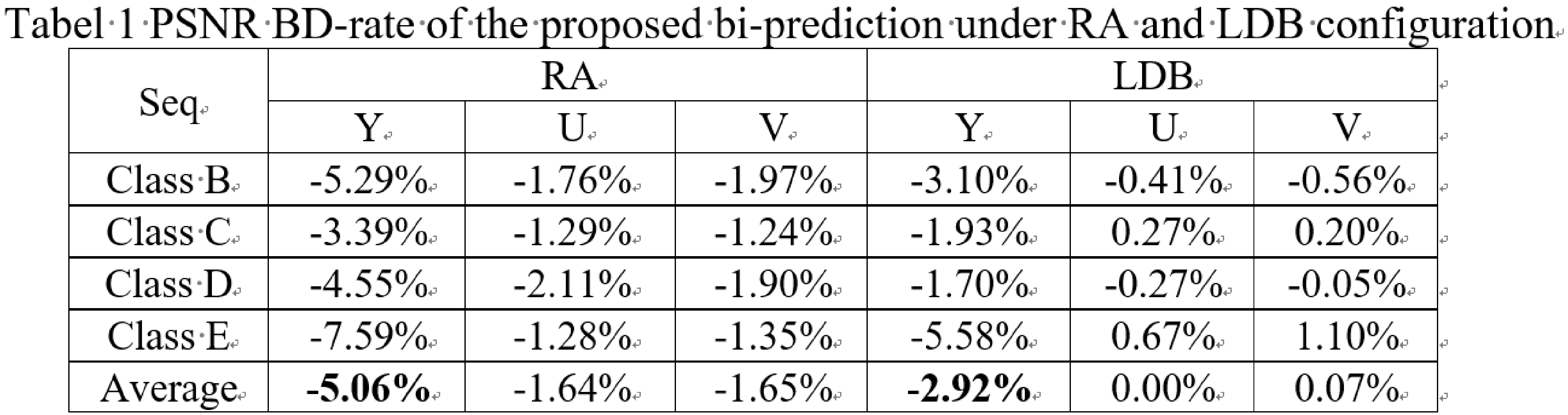
131차 MPEG DNNVC AHG: Input document 1 (2/3)

- Framework
  - To use spatial neighboring pixels
    - the reference blocks are extended toward left and upper
    - averaged bi-predictor are stitching with current blocks’ neighboring pixels
  - Picture order counts (POC) are used to measure temporal distances
  - A six-layer CNN model with skip connection



131차 MPEG DNNVC AHG: Input document 1 (3/3)

- Experimental results
  - By integrating the proposed CNN-based bi-prediction into HM16.15, it brings 2.92% and 5.06% bit-rate savings on average under Low-Delay B and Random Access configurations, respectively.

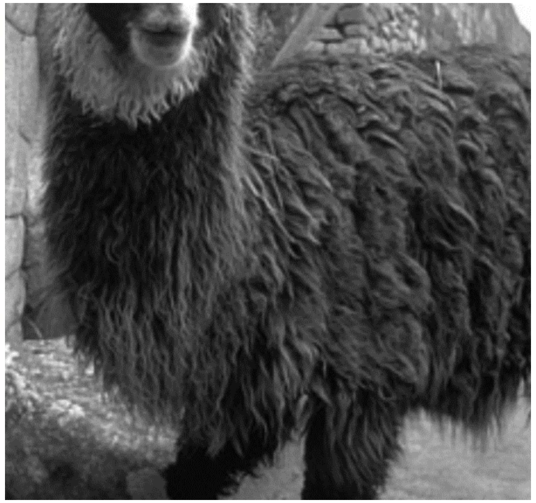


Tabel 1 PSNR BD-rate of the proposed bi-prediction under RA and LDB configuration

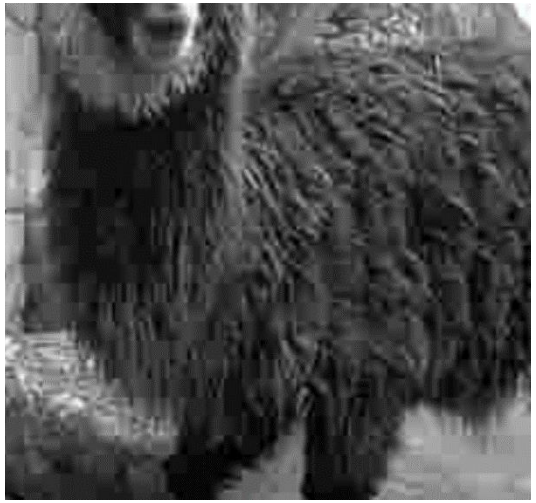
Seq	RA			LDB		
	Y	U	V	Y	U	V
Class B	-5.29%	-1.76%	-1.97%	-3.10%	-0.41%	-0.56%
Class C	-3.39%	-1.29%	-1.24%	-1.93%	0.27%	0.20%
Class D	-4.55%	-2.11%	-1.90%	-1.70%	-0.27%	-0.05%
Class E	-7.59%	-1.28%	-1.35%	-5.58%	0.67%	1.10%
Average	**-5.06%**	-1.64%	-1.65%	**-2.92%**	0.00%	0.07%

131차 MPEG DNNVC AHG: Input document 2 (1/4)

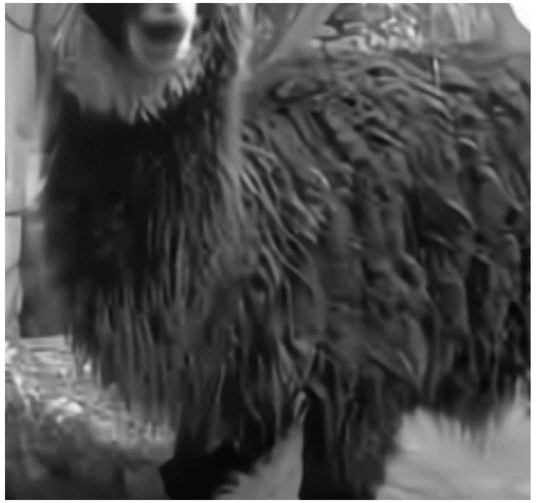
- [m54381] Learned Visual Quality Assessment for Image and Video Compression
  - Existing objective quality metrics that work well for conventional codecs maybe provide lower correlation with subjective evaluation for learning-based codecs.



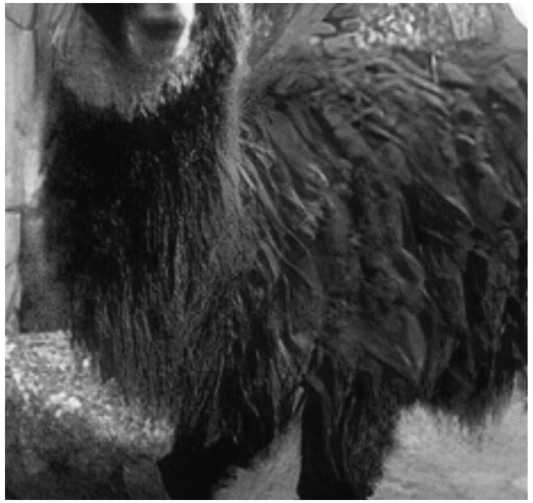
Original Image



JPEG (PSNR: 25.3dB)
Blocking



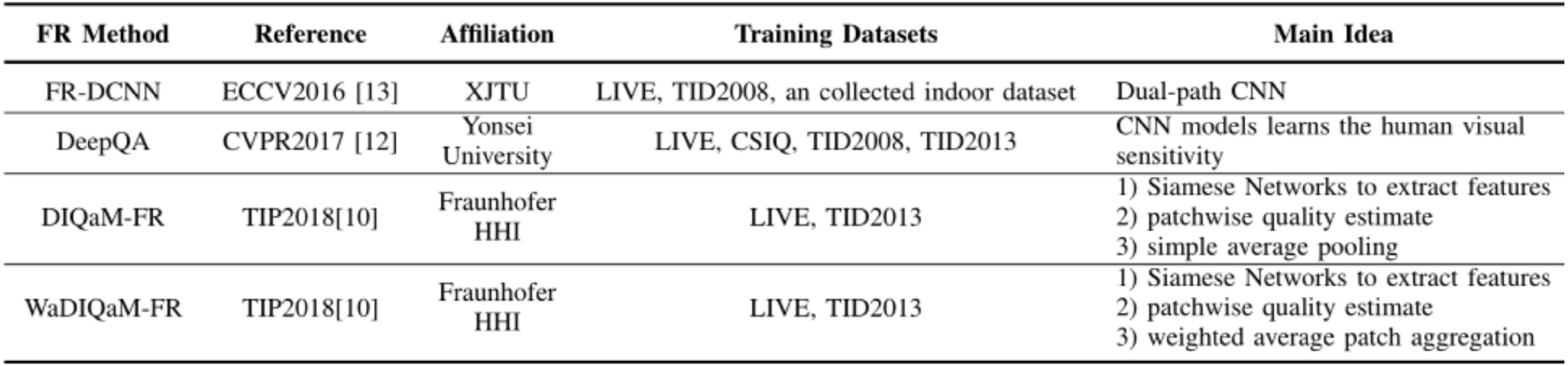
CNN (PSNR: 26.3dB)
More blurred



GAN (PSNR: 23.5dB)
Sharper

131차 MPEG DNNVC AHG: Input document 2 (2/4)

- Full Reference Image quality metrics based on DNN

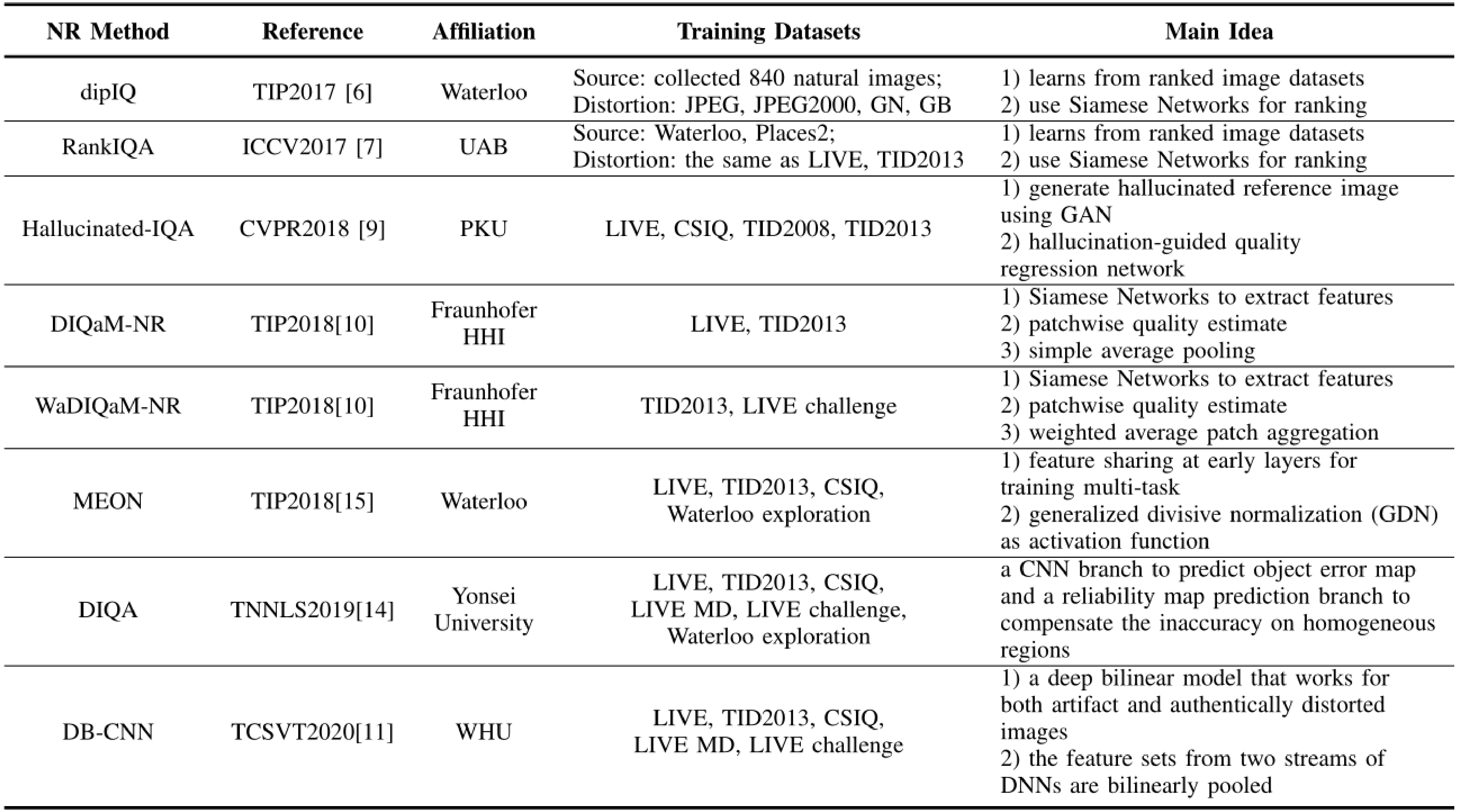


FR Method	Reference	Affiliation	Training Datasets	Main Idea
FR-DCNN	ECCV2016 [13]	XJTU	LIVE, TID2008, an collected indoor dataset	Dual-path CNN
DeepQA	CVPR2017 [12]	Yonsei University	LIVE, CSIQ, TID2008, TID2013	CNN models learns the human visual sensitivity
DIQaM-FR	TIP2018[10]	Fraunhofer HHI	LIVE, TID2013	1) Siamese Networks to extract features 2) patchwise quality estimate 3) simple average pooling
WaDIQaM-FR	TIP2018[10]	Fraunhofer HHI	LIVE, TID2013	1) Siamese Networks to extract features 2) patchwise quality estimate 3) weighted average patch aggregation

- LIVE: JPEG2000 compression, JPEG compression, white Gaussian noise, Gaussian blur and fast fading transmission error, 5~6 distortion levels
- CSIQ: JPEG2000 compression, JPEG compression, white Gaussian noise, Gaussian blur, pink Gaussian noise and global contrast decrements, 3~5 distortion levels
- TID2008: 17 distortions types, 4 distortion levels
- TID2013: 24 distortions types, 3~5 distortion levels

131차 MPEG DNNVC AHG: Input document 2 (3/4)

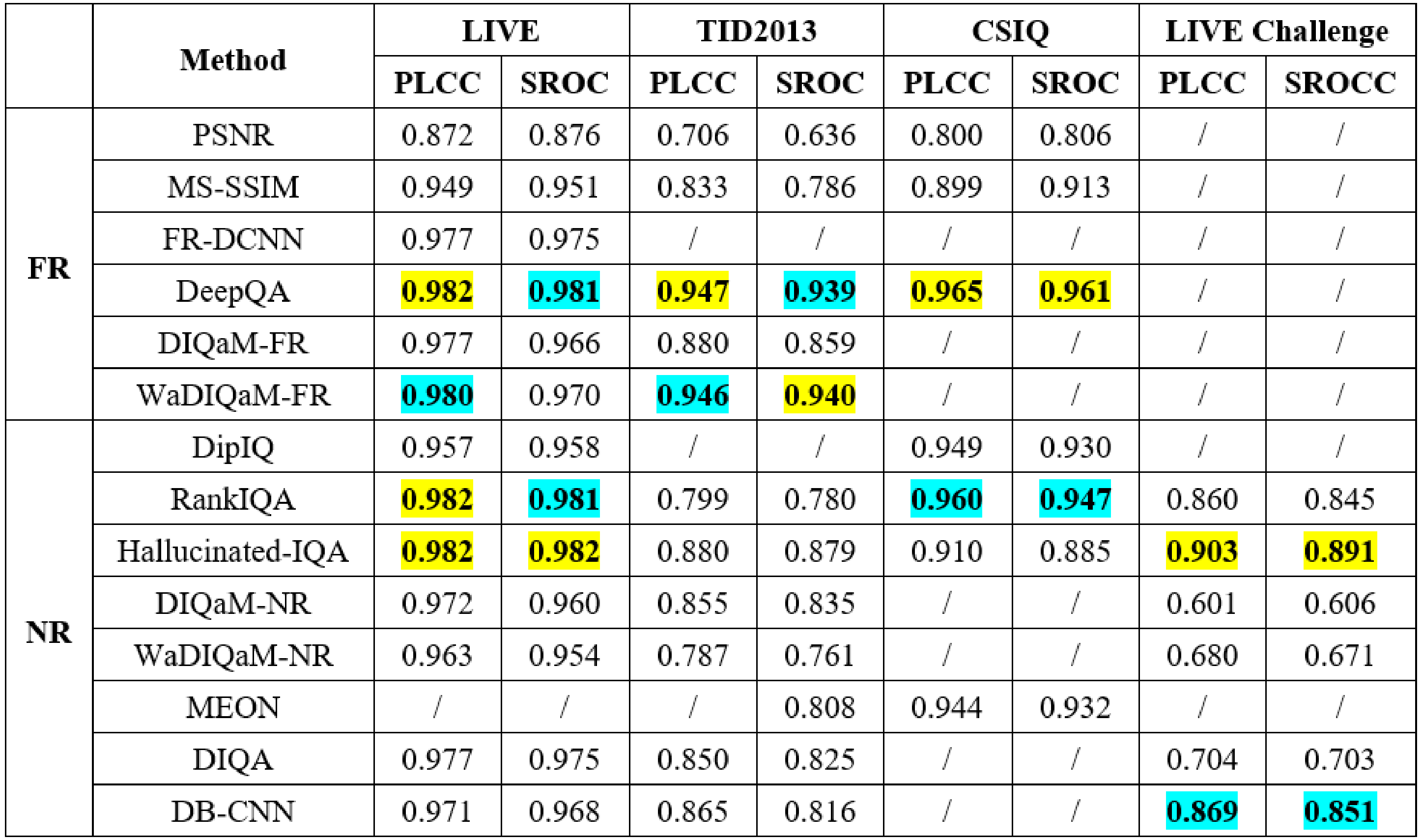
- No-Reference Image quality metrics based on DNN



NR Method	Reference	Affiliation	Training Datasets	Main Idea
dipIQ	TIP2017 [6]	Waterloo	Source: collected 840 natural images; Distortion: JPEG, JPEG2000, GN, GB	1) learns from ranked image datasets 2) use Siamese Networks for ranking
RankIQA	ICCV2017 [7]	UAB	Source: Waterloo, Places2; Distortion: the same as LIVE, TID2013	1) learns from ranked image datasets 2) use Siamese Networks for ranking
Hallucinated-IQA	CVPR2018 [9]	PKU	LIVE, CSIQ, TID2008, TID2013	1) generate hallucinated reference image using GAN 2) hallucination-guided quality regression network
DIQaM-NR	TIP2018[10]	Fraunhofer HHI	LIVE, TID2013	1) Siamese Networks to extract features 2) patchwise quality estimate 3) simple average pooling
WaDIQaM-NR	TIP2018[10]	Fraunhofer HHI	TID2013, LIVE challenge	1) Siamese Networks to extract features 2) patchwise quality estimate 3) weighted average patch aggregation
MEON	TIP2018[15]	Waterloo	LIVE, TID2013, CSIQ, Waterloo exploration	1) feature sharing at early layers for training multi-task 2) generalized divisive normalization (GDN) as activation function
DIQA	TNNLS2019[14]	Yonsei University	LIVE, TID2013, CSIQ, LIVE MD, LIVE challenge, Waterloo exploration	a CNN branch to predict object error map and a reliability map prediction branch to compensate the inaccuracy on homogeneous regions
DB-CNN	TCSVT2020[11]	WHU	LIVE, TID2013, CSIQ, LIVE MD, LIVE challenge	1) a deep bilinear model that works for both artifact and authentically distorted images 2) the feature sets from two streams of DNNs are bilinearly pooled

131차 MPEG DNNVC AHG: Input document 2 (4/4)

- Deep learning-based quality metrics Performances Report



	Method	LIVE		TID2013		CSIQ		LIVE Challenge	
		PLCC	SROC	PLCC	SROC	PLCC	SROC	PLCC	SROCC
FR	PSNR	0.872	0.876	0.706	0.636	0.800	0.806	/	/
	MS-SSIM	0.949	0.951	0.833	0.786	0.899	0.913	/	/
	FR-DCNN	0.977	0.975	/	/	/	/	/	/
	DeepQA	**0.982**	**0.981**	**0.947**	**0.939**	**0.965**	**0.961**	/	/
	DIQaM-FR	0.977	0.966	0.880	0.859	/	/	/	/
	WaDIQaM-FR	**0.980**	0.970	**0.946**	**0.940**	/	/	/	/
NR	DipIQ	0.957	0.958	/	/	0.949	0.930	/	/
	RankIQA	**0.982**	**0.981**	0.799	0.780	**0.960**	**0.947**	0.860	0.845
	Hallucinated-IQA	**0.982**	**0.982**	0.880	0.879	0.910	0.885	**0.903**	**0.891**
	DIQaM-NR	0.972	0.960	0.855	0.835	/	/	0.601	0.606
	WaDIQaM-NR	0.963	0.954	0.787	0.761	/	/	0.680	0.671
	MEON	/	/	/	0.808	0.944	0.932	/	/
	DIQA	0.977	0.975	0.850	0.825	/	/	0.704	0.703
	DB-CNN	0.971	0.968	0.865	0.816	/	/	**0.869**	**0.851**

131차 MPEG DNNVC AHG: Input document 3 (1/2)

- [m54467] A PyTorch library and evaluation platform for end-to-end compression research
  - Contributed by InterDigital Communications, Inc.
  - Authors: Jean Bégaint, Fabien Racapé, Simon Feltman, Akshay Pushparaja
  - Presents CompressAI, a platform that provides custom operations, layers, models and tools to research, develop and evaluate end-to-end image and video compression codecs.
- CompressAI
  - An open source library, available under the Apache license version 2.0
  - Publicly accessible on GitHub: https://github.com/InterDigitalInc/CompressAI
  - CompressAI aims to implement the most common operations to build deep neural network architectures for data compression in PyTorch, and to provide evaluation tools to compare learned methods against traditional codecs.
    - Also pre-defines SOTA model architectures with pre-trained weights.
    - Achieving similar performances as reported in the original papers

131차 MPEG DNNVC AHG: Input document 3 (2/2)

- CompressAI Models

Model name	Description	Related paper
bmshj2018_factorized	Factorized Prior model	J. Balle, D. Minnen, S. Singh, S.J. Hwang, N. Johnston: "Variational Image Compression with a Scale Hyperprior", (ICLR 2018)
bmshj2018_hyperprior	Scale Hyperprior model	
mbt2018_mean	Scale Hyperprior with non zero-mean Gaussian conditions	D. Minnen, J. Balle, G.D. Toderici: "Joint Autoregressive and Hierarchical Priors for Learned Image Compression", (NeurIPS 2018)
mbt2018	Joint Autoregressive Hierarchical Priors	
cheng2020_anchor*	Anchor model variant	Zhengxue Cheng, Heming Sun, Masaru Takeuchi, Jiro Katto: "Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules", (CVPR 2020)
cheng2020_attn*	Self-attention model variant	

*: Pre-trained weights are not yet available

GISC2020
Global ICT Standards Conference

뉴 노멀 시대
선도를 위한
ICT 표준의
역할

131차 MPEG DNNVC BoG

- Issue
  - Data set and rights to use
    - Collect a list of data sets
  - Use case and requirements
    - Start the drafting
    - Start from VVC use case and requirements documents
  - Test conditions and Quality metrics
- Output Document
  - Draft test conditions for DNNVC (w19583)
  - Draft requirements and use case for DNNVC (w19582)

Draft test conditions for DNNVC (1/3)

- Test sequence
  - VVC YUV420 CTC sequence
    - A1-F: mandatory
    - H3 (HDR): optional
  - length: all frame except AI
    - AI: the first of every 8 frames
  - Randomized test set (further study)
    - to avoid overfitting

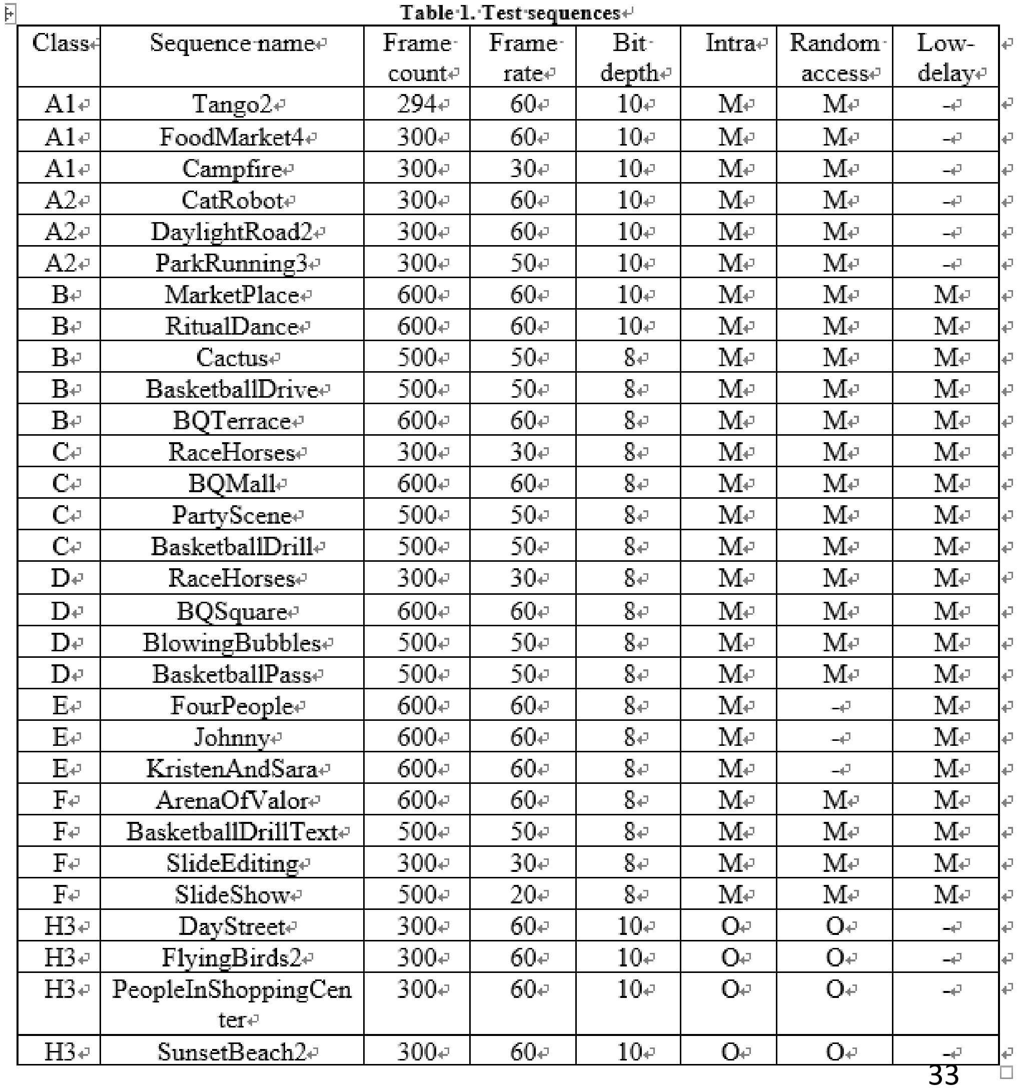


Table 1. Test sequences

Class	Sequence name	Frame count	Frame rate	Bit depth	Intra	Random access	Low-delay
A1	Tango2	294	60	10	M	M	-
A1	FoodMarket4	300	60	10	M	M	-
A1	Campfire	300	30	10	M	M	-
A2	CatRobot	300	60	10	M	M	-
A2	DaylightRoad2	300	60	10	M	M	-
A2	ParkRunning3	300	50	10	M	M	-
B	MarketPlace	600	60	10	M	M	M
B	RitualDance	600	60	10	M	M	M
B	Cactus	500	50	8	M	M	M
B	BasketballDrive	500	50	8	M	M	M
B	BQTerrace	600	60	8	M	M	M
C	RaceHorses	300	30	8	M	M	M
C	BQMall	600	60	8	M	M	M
C	PartyScene	500	50	8	M	M	M
C	BasketballDrill	500	50	8	M	M	M
D	RaceHorses	300	30	8	M	M	M
D	BQSquare	600	60	8	M	M	M
D	BlowingBubbles	500	50	8	M	M	M
D	BasketballPass	500	50	8	M	M	M
E	FourPeople	600	60	8	M	-	M
E	Johnny	600	60	8	M	-	M
E	KristenAndSara	600	60	8	M	-	M
F	ArenaOfValor	600	60	8	M	M	M
F	BasketballDrillText	500	50	8	M	M	M
F	SlideEditing	300	30	8	M	M	M
F	SlideShow	500	20	8	M	M	M
H3	DayStreet	300	60	10	O	O	-
H3	FlyingBirds2	300	60	10	O	O	-
H3	PeopleInShoppingCenter	300	60	10	O	O	-
H3	SunsetBeach2	300	60	10	O	O	-

Draft test conditions for DNNVC (2/3)

- Coding Condition
  - AI, RA, LD
  - Bit depth: 10 bit (internal conversion)
  - Bit rate
    - Fixed QP: [22, 27, 32, 37], [27, 32, 37, 42], etc
    - Bit-rate target: 4 RD points, anchor QP ± 10%
- Anchor
  - VTM 10.0
  - should be reported if parallel encoding/decoding, (JVET-B0036) for RA
    - report precise 64-bit PSNR value (w/ PrintHexPSNR)

Draft test conditions for DNNVC (3/3)

- Training
  - training sequences are under discussion
- Performance Report
  - Metrics: PSNR, MS-SSIM
  - Complexity: anchor and proposed encoding and decoding running times @similar CPU & GPU configuration
  - Optional:
    - test environment
    - network environment – loss function, optimizer, batch size, etc
    - network information – the size of the network, parameter precision, etc
    - framework – Pytorch, TensorFlow, etc

목차

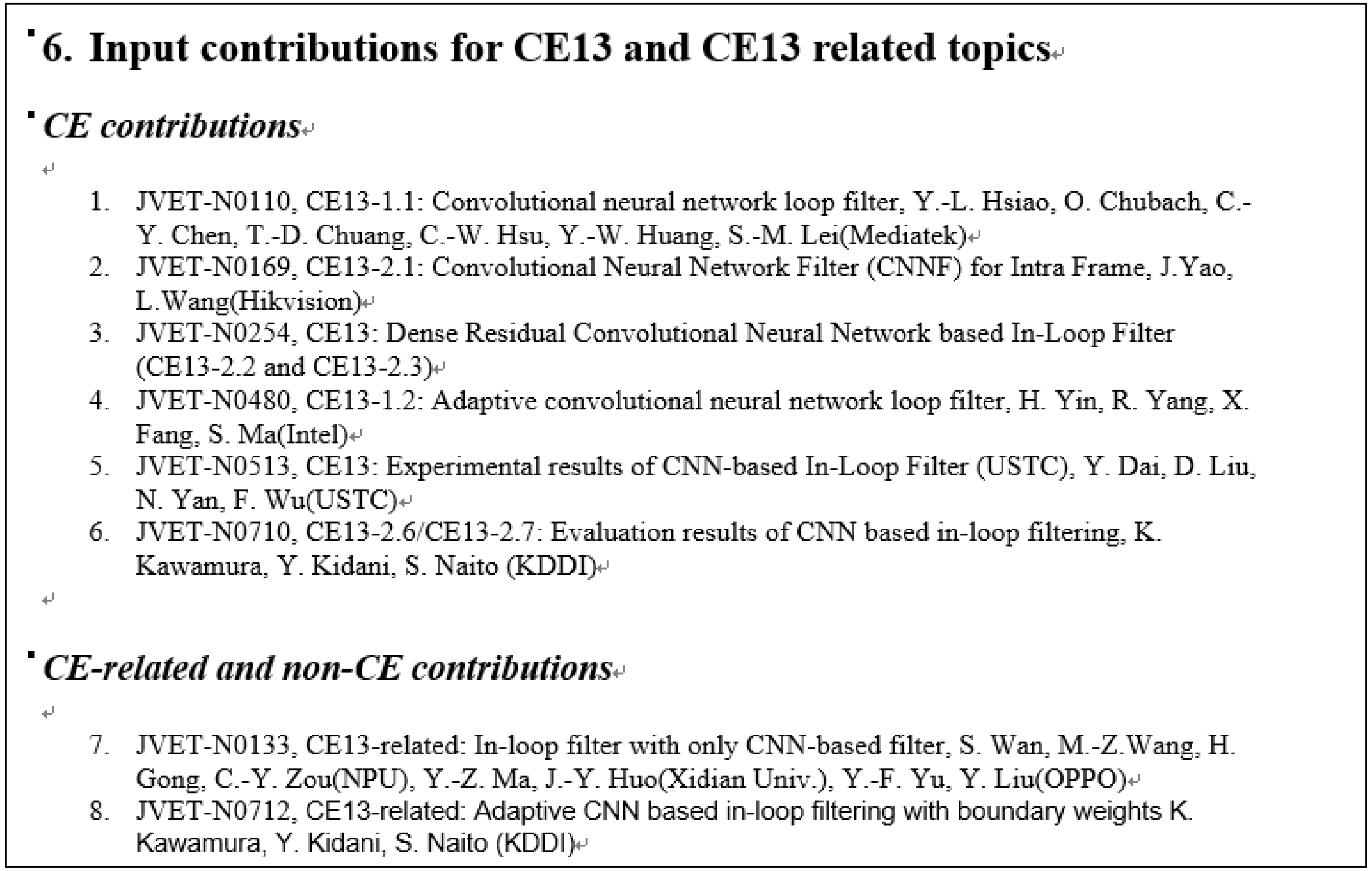
- Introduction
- NN-base video coding approaches
- MPEG DNNVC
- JVET NNVC
- Conclusion

JVET AHG11 NNVC

- History: AHG9
  - JVET AHG9 (neural network tool in video coding) was set up for investigating the compression performance and complexity of neural network-based coding tools in early 2018
  - AHG9 Mandates
    - Investigate the benefit of deep learning technology in video compression.
    - Investigate the complexity impact of using deep learning in video compression.
    - Investigate deep learning based coding tools such as CNN loop filter.
    - Investigate the relationship between CNN filter and ALF, and other loop filters.
    - Investigate the performance of CNN filter used as an in-loop filter or a post-processing filter.
    - Investigate the impact of QP on CNN filter.
  - ended after the P meeting (Oct 2019)
  - 2 output documents: JVET-L1006 and JVET-M1006
  - Core Experiments results: 4.6% gain (AI, Y) over VTM-4.0, nearly 4% gain (AI, Y) over VTM-5.0. Higher coding gains were observed for chroma components in general.

JVET AHG11 NNVC

- History: AHG9
  - CE13: Summary Report on Neural Network based Filter for Video Coding, JVET-N0033



6. Input contributions for CE13 and CE13 related topics↵

CE contributions↵

↵

1. JVET-N0110, CE13-1.1: Convolutional neural network loop filter, Y.-L. Hsiao, O. Chubach, C.-Y. Chen, T.-D. Chuang, C.-W. Hsu, Y.-W. Huang, S.-M. Lei(Mediatek)↵
2. JVET-N0169, CE13-2.1: Convolutional Neural Network Filter (CNNF) for Intra Frame, J.Yao, L.Wang(Hikvision)↵
3. JVET-N0254, CE13: Dense Residual Convolutional Neural Network based In-Loop Filter (CE13-2.2 and CE13-2.3)↵
4. JVET-N0480, CE13-1.2: Adaptive convolutional neural network loop filter, H. Yin, R. Yang, X. Fang, S. Ma(Intel)↵
5. JVET-N0513, CE13: Experimental results of CNN-based In-Loop Filter (USTC), Y. Dai, D. Liu, N. Yan, F. Wu(USTC)↵
6. JVET-N0710, CE13-2.6/CE13-2.7: Evaluation results of CNN based in-loop filtering, K. Kawamura, Y. Kidani, S. Naito (KDDI)↵

↵

CE-related and non-CE contributions↵

↵

7. JVET-N0133, CE13-related: In-loop filter with only CNN-based filter, S. Wan, M.-Z.Wang, H. Gong, C.-Y. Zou(NPU), Y.-Z. Ma, J.-Y. Huo(Xidian Univ.), Y.-F. Yu, Y. Liu(OPPO)↵
8. JVET-N0712, CE13-related: Adaptive CNN based in-loop filtering with boundary weights K. Kawamura, Y. Kidani, S. Naito (KDDI)↵

JVET AHG11 NNVC

- History: AHG9
  - CE13: Summary Report on Neural Network based Filter for Video Coding, JVET-N0033

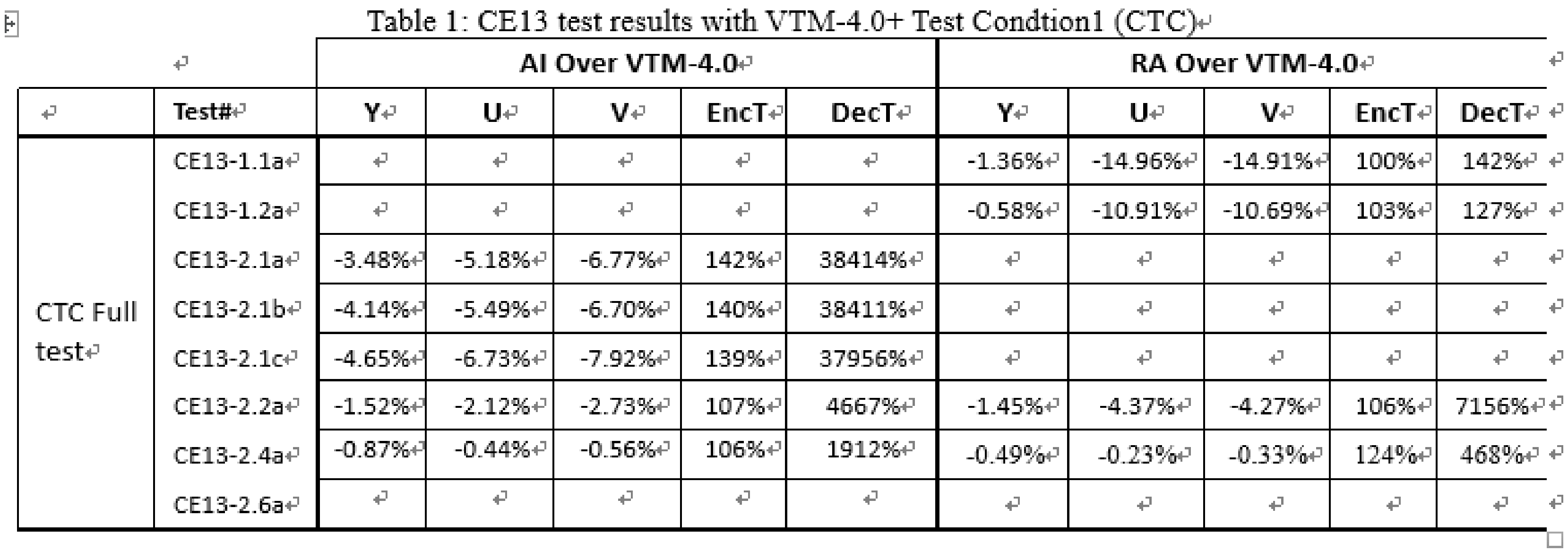
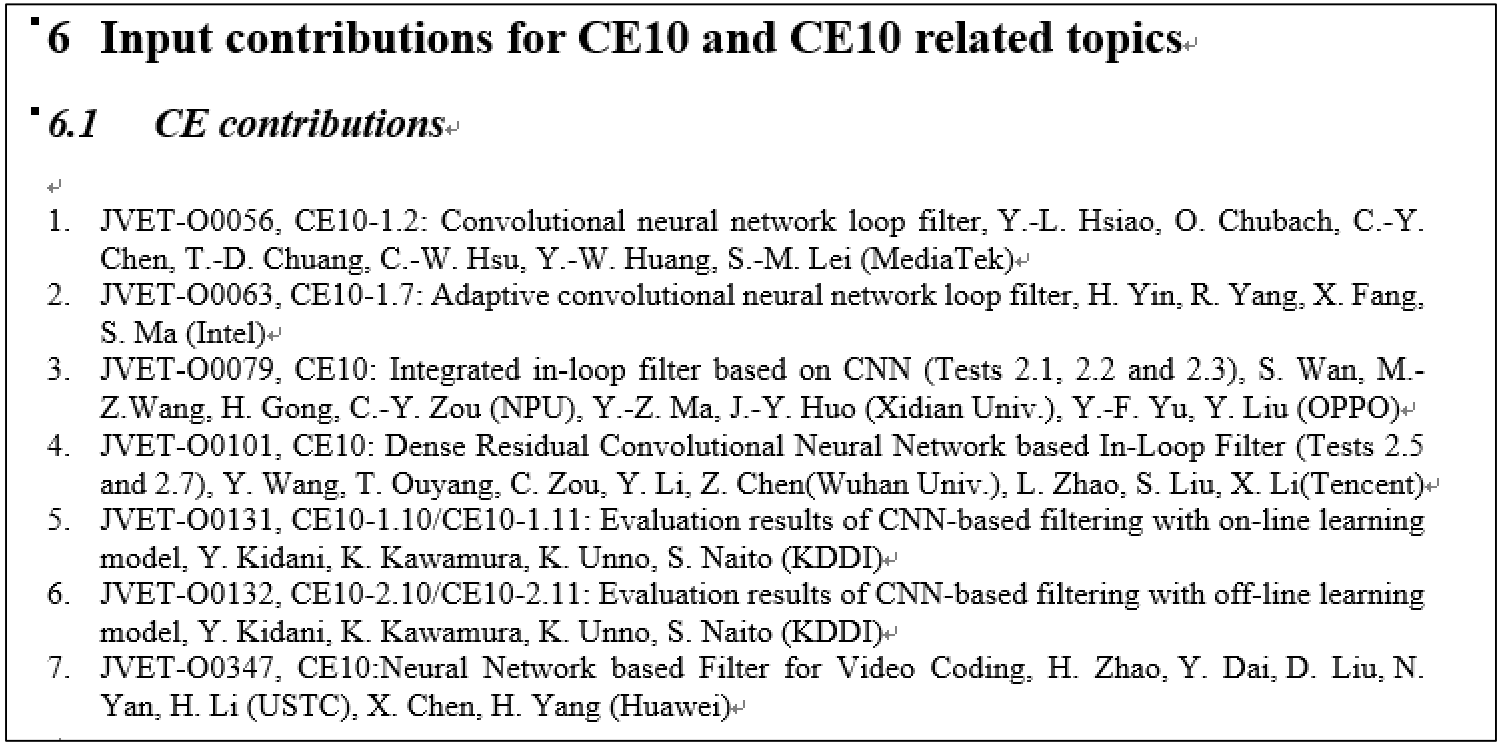


Table 1: CE13 test results with VTM-4.0+ Test Condtion1 (CTC)

		AI Over VTM-4.0					RA Over VTM-4.0				
	Test#	Y	U	V	EncT	DecT	Y	U	V	EncT	DecT
CTC Full test	CE13-1.1a						-1.36%	-14.96%	-14.91%	100%	142%
	CE13-1.2a						-0.58%	-10.91%	-10.69%	103%	127%
	CE13-2.1a	-3.48%	-5.18%	-6.77%	142%	38414%					
	CE13-2.1b	-4.14%	-5.49%	-6.70%	140%	38411%					
	CE13-2.1c	-4.65%	-6.73%	-7.92%	139%	37956%					
	CE13-2.2a	-1.52%	-2.12%	-2.73%	107%	4667%	-1.45%	-4.37%	-4.27%	106%	7156%
	CE13-2.4a	-0.87%	-0.44%	-0.56%	106%	1912%	-0.49%	-0.23%	-0.33%	124%	468%
	CE13-2.6a										

JVET AHG11 NNVC

- History: AHG9
  - CE10: Summary Report on Neural Network based Filter for Video Coding, JVET-O0030



6 Input contributions for CE10 and CE10 related topics

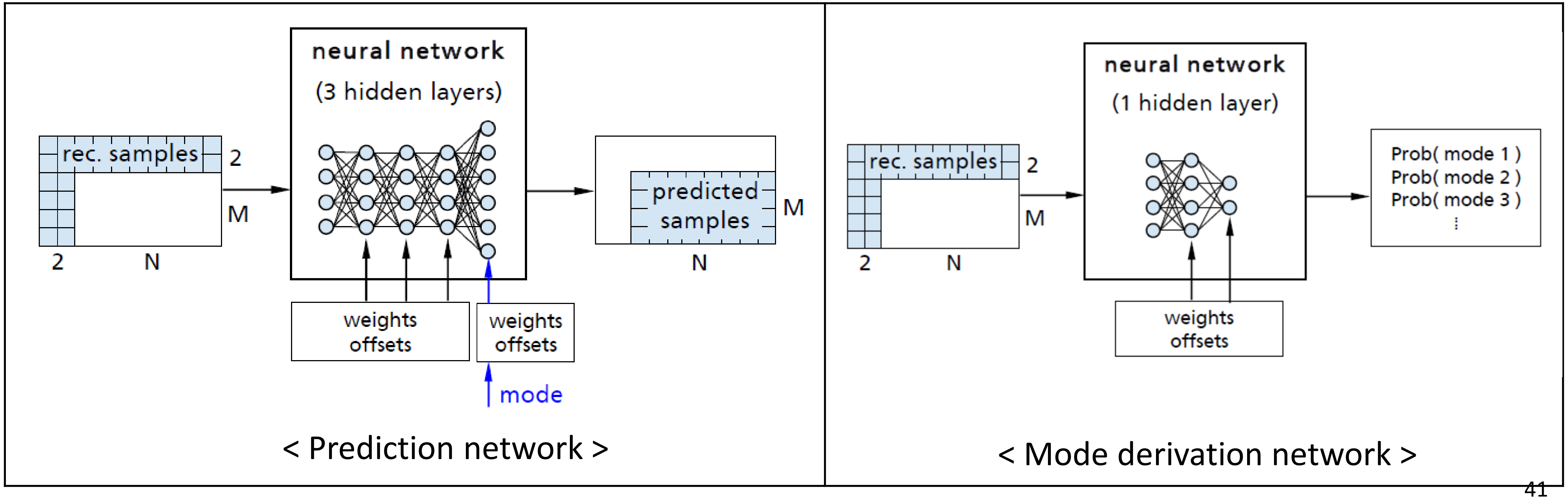
6.1 CE contributions

1. JVET-O0056, CE10-1.2: Convolutional neural network loop filter, Y.-L. Hsiao, O. Chubach, C.-Y. Chen, T.-D. Chuang, C.-W. Hsu, Y.-W. Huang, S.-M. Lei (MediaTek)
2. JVET-O0063, CE10-1.7: Adaptive convolutional neural network loop filter, H. Yin, R. Yang, X. Fang, S. Ma (Intel)
3. JVET-O0079, CE10: Integrated in-loop filter based on CNN (Tests 2.1, 2.2 and 2.3), S. Wan, M.-Z.Wang, H. Gong, C.-Y. Zou (NPU), Y.-Z. Ma, J.-Y. Huo (Xidian Univ.), Y.-F. Yu, Y. Liu (OPPO)
4. JVET-O0101, CE10: Dense Residual Convolutional Neural Network based In-Loop Filter (Tests 2.5 and 2.7), Y. Wang, T. Ouyang, C. Zou, Y. Li, Z. Chen(Wuhan Univ.), L. Zhao, S. Liu, X. Li(Tencent)
5. JVET-O0131, CE10-1.10/CE10-1.11: Evaluation results of CNN-based filtering with on-line learning model, Y. Kidani, K. Kawamura, K. Unno, S. Naito (KDDI)
6. JVET-O0132, CE10-2.10/CE10-2.11: Evaluation results of CNN-based filtering with off-line learning model, Y. Kidani, K. Kawamura, K. Unno, S. Naito (KDDI)
7. JVET-O0347, CE10:Neural Network based Filter for Video Coding, H. Zhao, Y. Dai, D. Liu, N. Yan, H. Li (USTC), X. Chen, H. Yang (Huawei)

JVET AHG11 NNVC

- History: VVC MIP (Matrix Weighted Intra Prediction)
  - "Intra prediction using neural networks," JVET-J0014, April 2018
    - Prediction network
    - Mode derivation network

RA: -1.36% gain w/ 124% enc time, 104% dec time



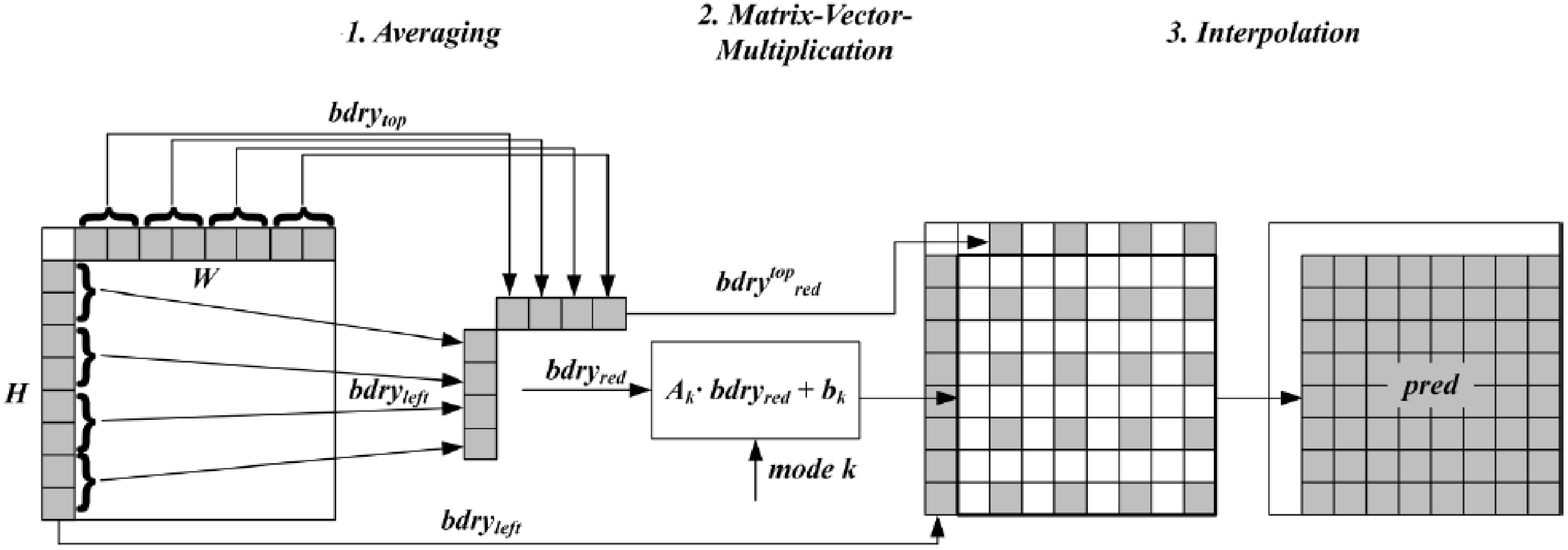
< Prediction network >

< Mode derivation network >

JVET AHG11 NNVC

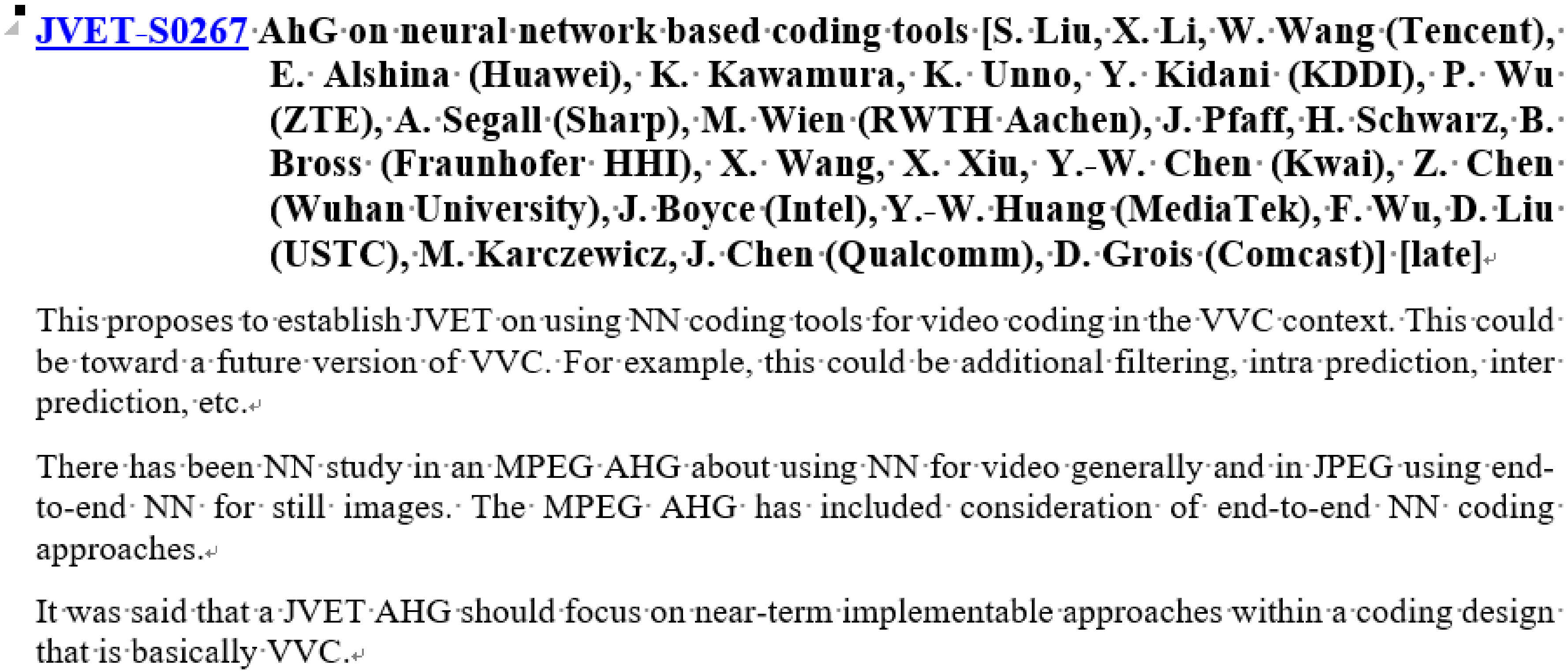
- History: VVC MIP (Matrix Weighted Intra Prediction)
  - “Affine Linear Weighted Intra prediction,” JVET-N0217, March 2019
    - Averaging
    - Matrix-Vector Multiplication
    - Interpolation

RA: -0.42% gain w/ 104% enc time, 99% dec time



JVET AHG11 NNVC

- JVET AHG 11 (Proposal JVET-S0267)



JVET-S0267 AhG on neural network based coding tools [S. Liu, X. Li, W. Wang (Tencent), E. Alshina (Huawei), K. Kawamura, K. Unno, Y. Kidani (KDDI), P. Wu (ZTE), A. Segall (Sharp), M. Wien (RWTH Aachen), J. Pfaff, H. Schwarz, B. Bross (Fraunhofer HHI), X. Wang, X. Xiu, Y.-W. Chen (Kwai), Z. Chen (Wuhan University), J. Boyce (Intel), Y.-W. Huang (MediaTek), F. Wu, D. Liu (USTC), M. Karczewicz, J. Chen (Qualcomm), D. Grois (Comcast)] [late]

This proposes to establish JVET on using NN coding tools for video coding in the VVC context. This could be toward a future version of VVC. For example, this could be additional filtering, intra prediction, inter prediction, etc.

There has been NN study in an MPEG AHG about using NN for video generally and in JPEG using end-to-end NN for still images. The MPEG AHG has included consideration of end-to-end NN coding approaches.

It was said that a JVET AHG should focus on near-term implementable approaches within a coding design that is basically VVC.

JVET AHG11 NNVC

- JVET AHG 11 (Proposal JVET-S0267)
  - Mandates
    - Study potential extensions of VVC with NN-based coding tools for video coding, such as intra or inter prediction modes, partitioning, transforms, and in-loop or post filtering.
    - Study NN-based encoding optimization for VVC.
    - Study the impact of training on the performance of candidate technology.
    - Analyze complexity characteristics and perform complexity analysis of candidate technology.
    - Identify video test materials, training set materials, and testing methods for assessment of the effectiveness and complexity of considered tools.
    - Develop reporting templates for test results and analysis of candidate technology.
    - Coordinate with relevant activities of the parent bodies.
  - Chairmen: E. Alshina, S. Liu, J. Pfaff, M. Wien, P. Wu, Y. Ye (co-chairs)

JVET AHG11 NNVC

- First teleconference meeting
  - July 30, 14:00-16:00 UTC
  - Input document: JVET-T0041
    - 1) Methodology and materials for training
    - 2) Methodology and materials for inference and testing
    - 3) Anchor and reporting template
  - Output document: JVET-T0042

JVET AHG11 NNVC

- 1) Methodology for training
  - Training data
    - Training data are under investigation, copyright check
    - Color Encoding: YUV
    - Color Sampling: 4:2:0
    - Color Range: Limited
    - Bit-depth: 10 bits
    - Dimensions: Even Height and Width
  - Training principles
    - use common pre-identified training sequence sets
    - generate training set using the recent VTM 10.0
    - describe training principles and provide a training example in text
  - Complexity of training stage
    - HW/SW environment (e.g., CPU, GPU, OS)
    - Framework (e.g., Pytorch)
    - Epoch, Batch size, Training time
    - Additionally patch size, learning rate, optimizer, loss function, preprocessing

JVET AHG11 NNVC

- 2) Methodology for inference
  - Testing data
    - BD-rate using QP {22, 27, 32, 27, 42} @ JVET-T0041-ReportingTemplate.xlsm
  - Complexity of inference stage
    - Total Conv. layers, FC layers, Framework
    - Param. Number, Param. Precision
    - Memory usage (MB) of total param.
    - Temporary memory (MB) to store output feature map
  - Study MS-SSIM and other quality metrics (AHG4)
  - Compare Low Delay 4K performance between HEVC and VVC

JVET AHG11 NNVC

- 3) Reporting template: training stage

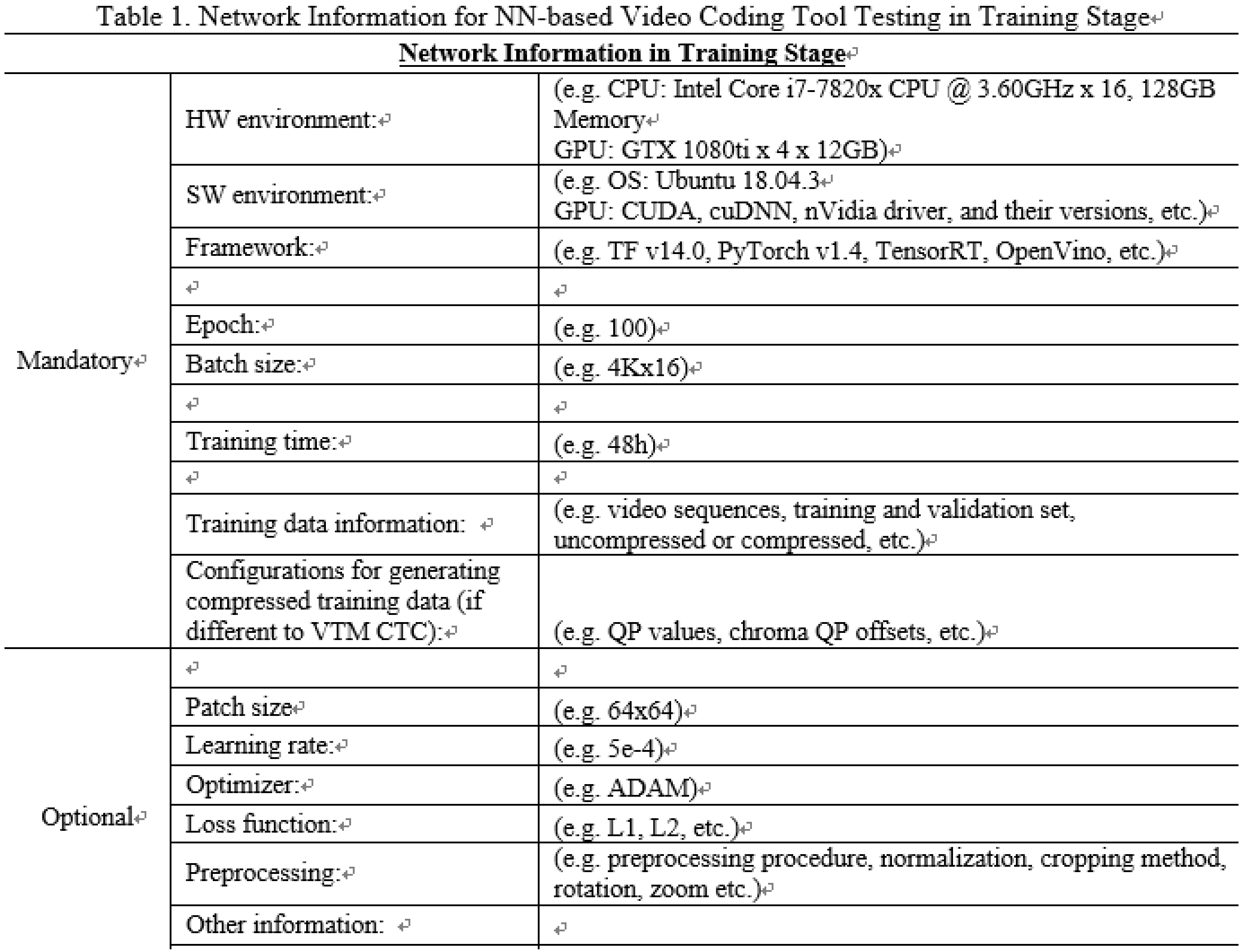


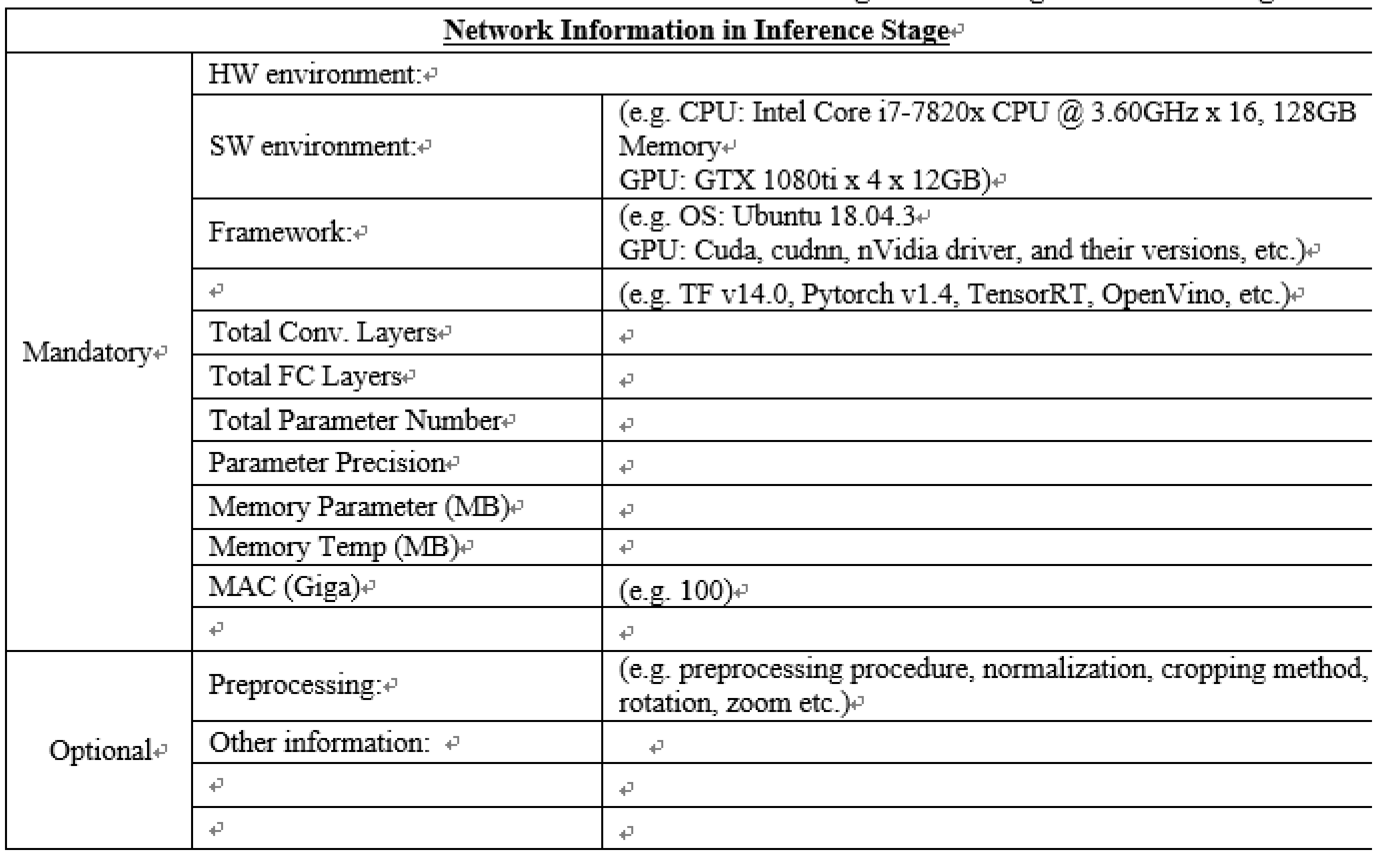
Table 1. Network Information for NN-based Video Coding Tool Testing in Training Stage

Network Information in Training Stage		
Mandatory	HW environment:	(e.g. CPU: Intel Core i7-7820x CPU @ 3.60GHz x 16, 128GB Memory GPU: GTX 1080ti x 4 x 12GB)
	SW environment:	(e.g. OS: Ubuntu 18.04.3 GPU: CUDA, cuDNN, nVidia driver, and their versions, etc.)
	Framework:	(e.g. TF v14.0, PyTorch v1.4, TensorRT, OpenVino, etc.)
	Epoch:	(e.g. 100)
	Batch size:	(e.g. 4Kx16)
	Training time:	(e.g. 48h)
	Training data information:	(e.g. video sequences, training and validation set, uncompressed or compressed, etc.)
	Configurations for generating compressed training data (if different to VTM CTC):	(e.g. QP values, chroma QP offsets, etc.)
Optional		
	Patch size	(e.g. 64x64)
	Learning rate:	(e.g. 5e-4)
	Optimizer:	(e.g. ADAM)
	Loss function:	(e.g. L1, L2, etc.)
	Preprocessing:	(e.g. preprocessing procedure, normalization, cropping method, rotation, zoom etc.)
	Other information:	

JVET AHG11 NNVC

- 3) Reporting template: testing stage

Table 2. Network Information for NN-based Video Coding Tool Testing in Inference Stage



Network Information in Inference Stage		
Mandatory	HW environment:	
	SW environment:	(e.g. CPU: Intel Core i7-7820x CPU @ 3.60GHz x 16, 128GB Memory GPU: GTX 1080ti x 4 x 12GB)
	Framework:	(e.g. OS: Ubuntu 18.04.3 GPU: Cuda, cudnn, nVidia driver, and their versions, etc.)
		(e.g. TF v14.0, Pytorch v1.4, TensorRT, OpenVino, etc.)
	Total Conv. Layers	
	Total FC Layers	
	Total Parameter Number	
	Parameter Precision	
	Memory Parameter (MB)	
	Memory Temp (MB)	
	MAC (Giga)	(e.g. 100)
Optional	Preprocessing:	(e.g. preprocessing procedure, normalization, cropping method, rotation, zoom etc.)
	Other information:	

Conclusion

- Significant advances in video coding can be observed by the development of deep neural networks
- Neural networks in hybrid coding framework and end-to-end optimized coding
- MPEG DNNVC and JVET NNVC
- Compression gain and Complexity

Thank You

kimyounhee@etri.re.kr