Learned Visual Data Compression Technologies and Standards

Dr. Shan Liu
Distinguished Scientist and General Manager
Tencent Media Lab
July 2021

Outline

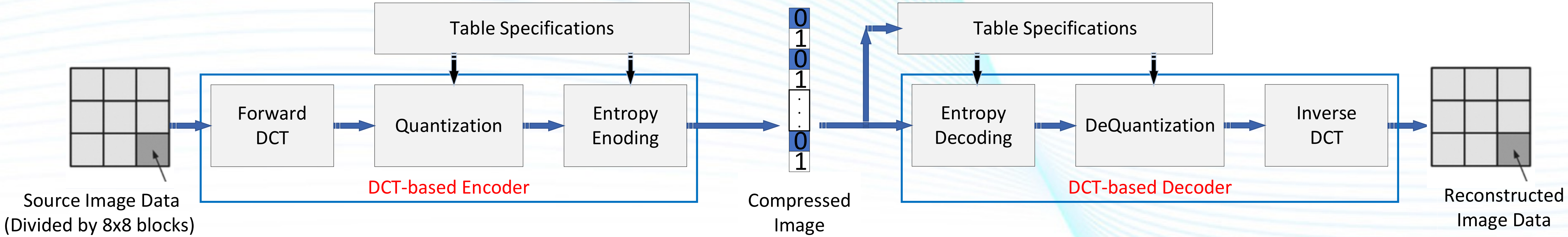
- Overview of learning-based visual data compression
  - Learning-based image compression
  - Learning-based video compression
  - Learning-based volumetric visual data compression
- Standard activities on learning-based visual data compression
  - JPEG AI
  - JVET NNVC
  - IEEE DCSC FVC
- Discussion and related work

Outline

- Overview of learning-based visual data compression
  - Learning-based image compression
  - Learning-based video compression
  - Learning-based volumetric visual data compression
- Standard activities on learning-based visual data compression
  - JPEG AI
  - JVET NNVC
  - IEEE DCSC FVC
- Discussion and related work

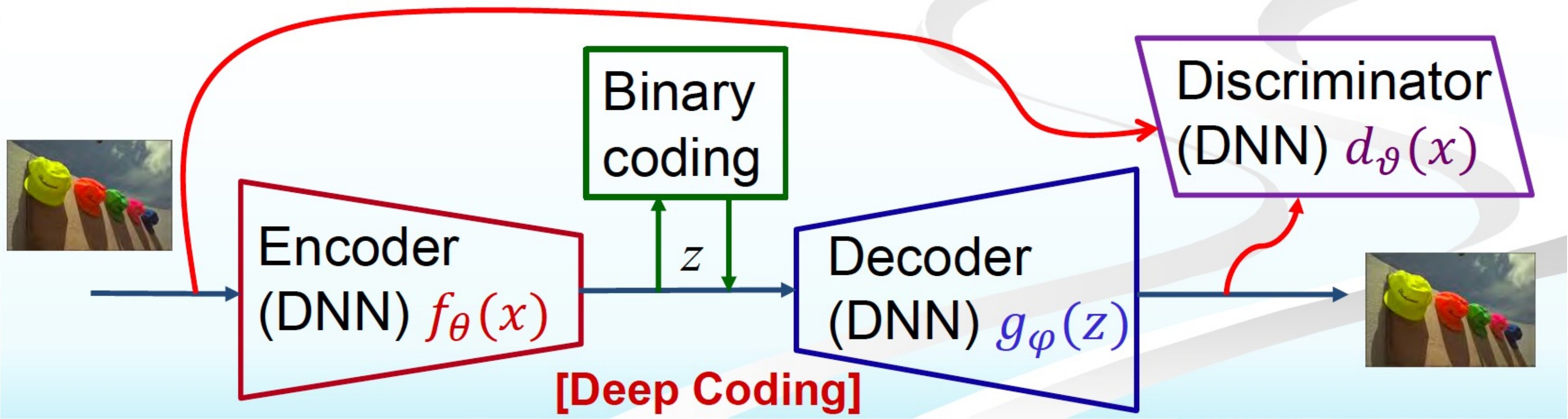
Conventional Image Compression

- Framework



- Standards: JPEG (DCT + Huffman), JPEG2000 (DWT + AC), BPG (HEVC intra)

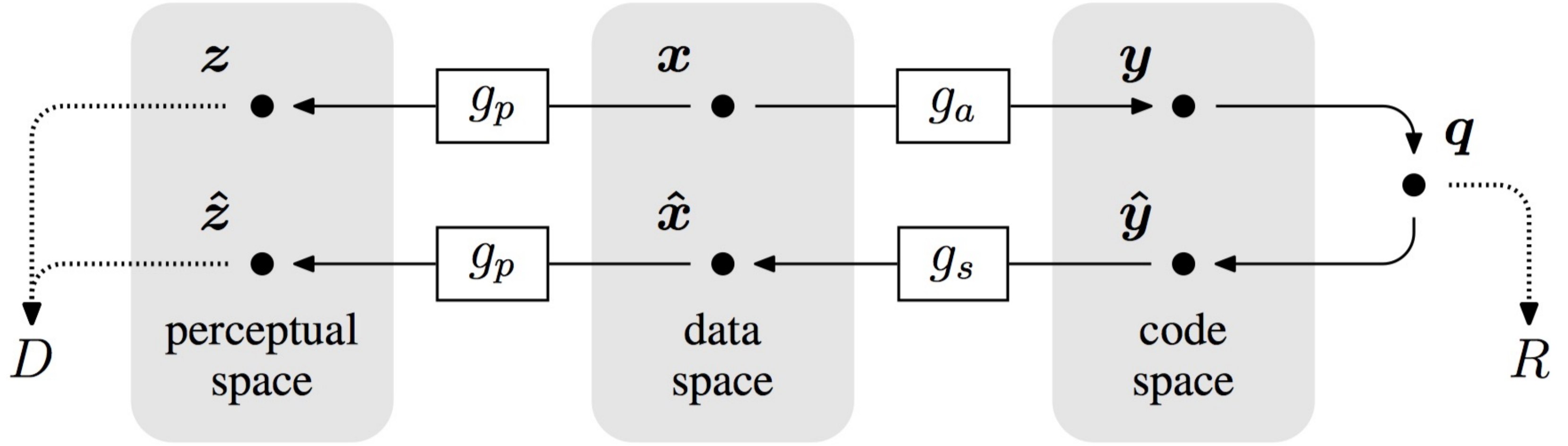
Learned Image Compression



- Learning-assisted image compression
- Hybrid learned image compression
- End-to-end learned image compression (Autoencoder)

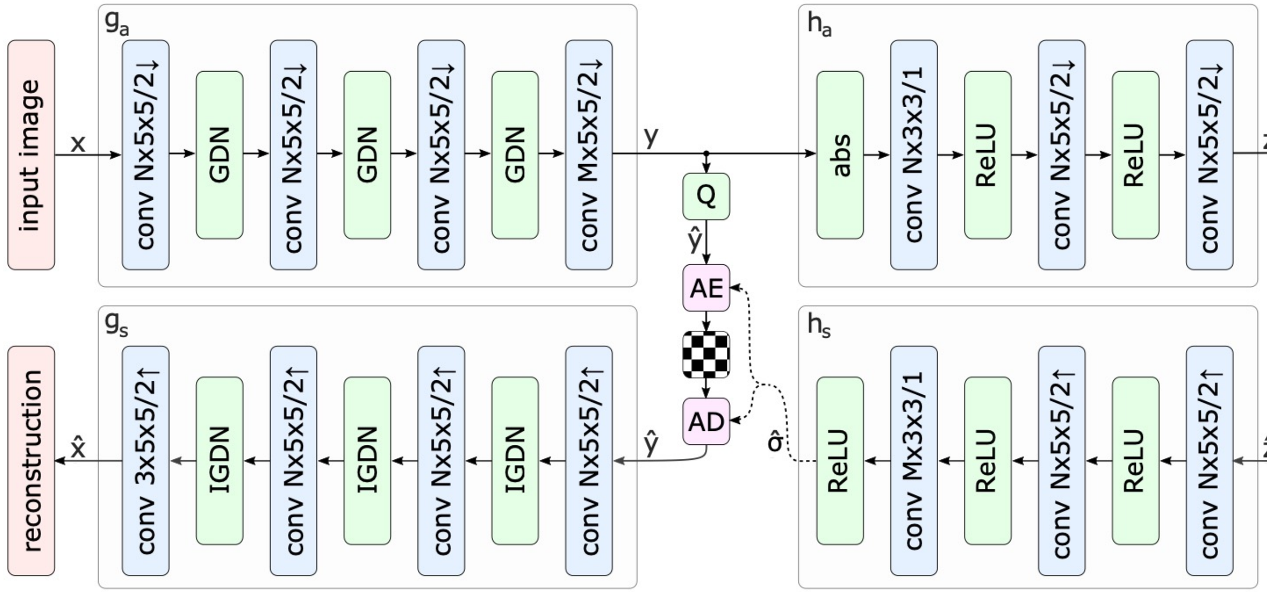
E2E Learned Image Compression

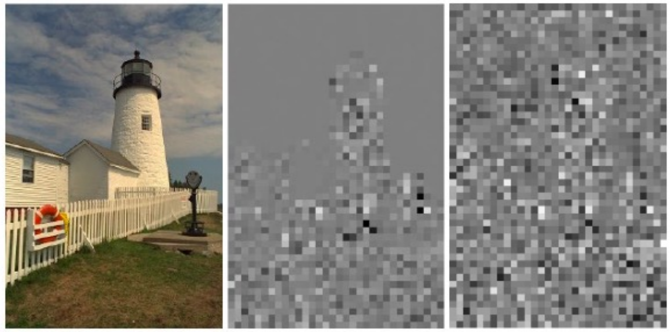
- Basic Framework



Ballé, Johannes, et al. "End-to-end optimized image compression", ICLR 2017.

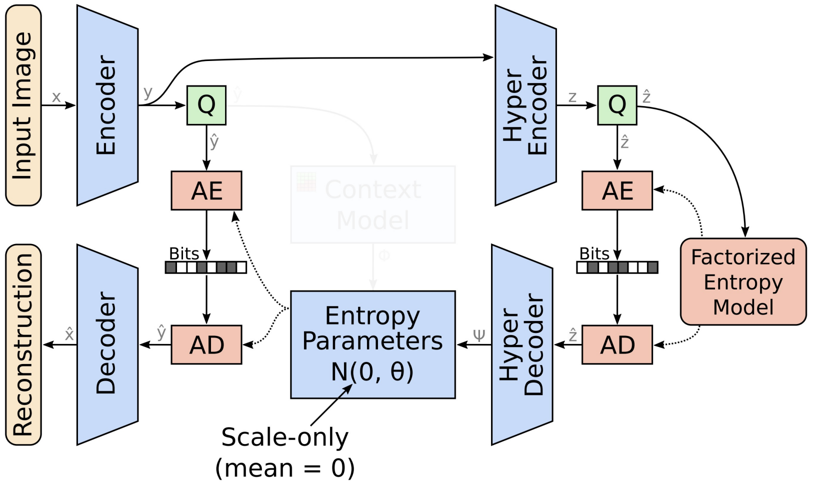
E2E Learned Image Compression



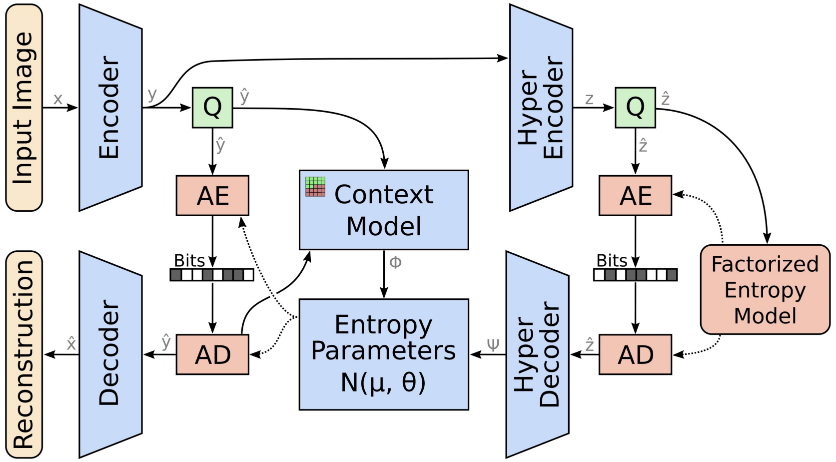


Ballé, Johannes, et al. "Variational image compression with a scale hyperprior", ICLR 2018.

Minnen, David, et al. "Joint autoregressive and hierarchical priors for learned image compression", in NeruIPS 2018.

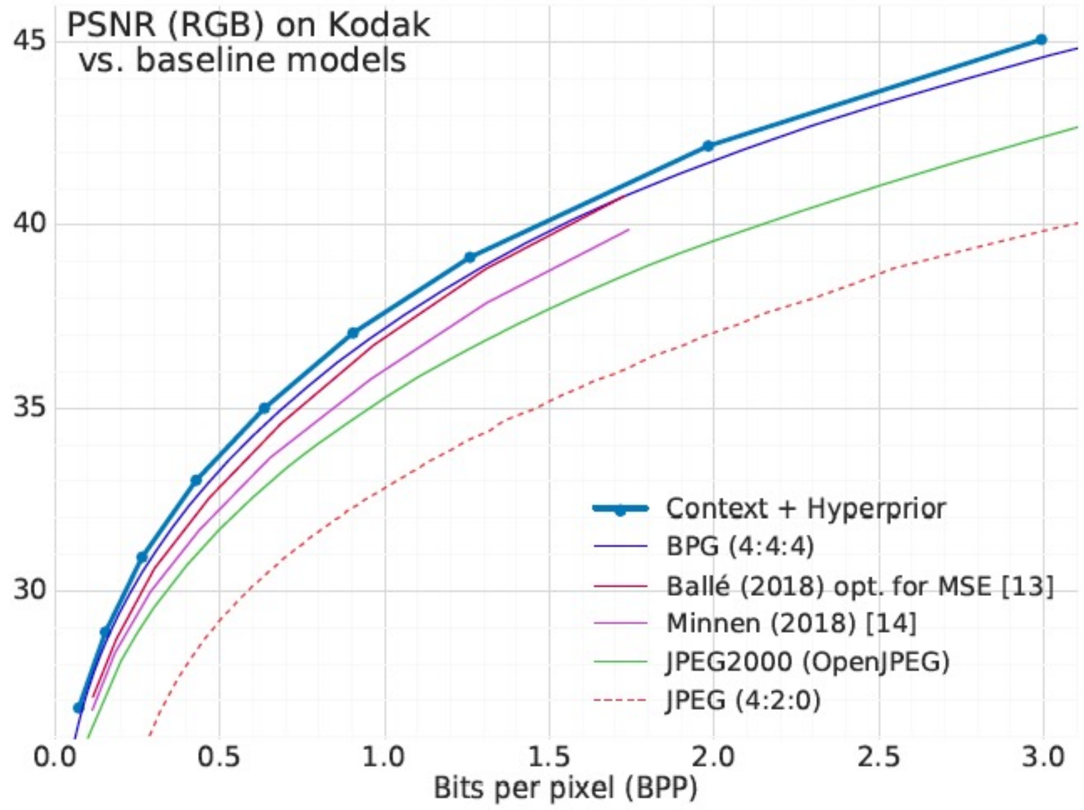


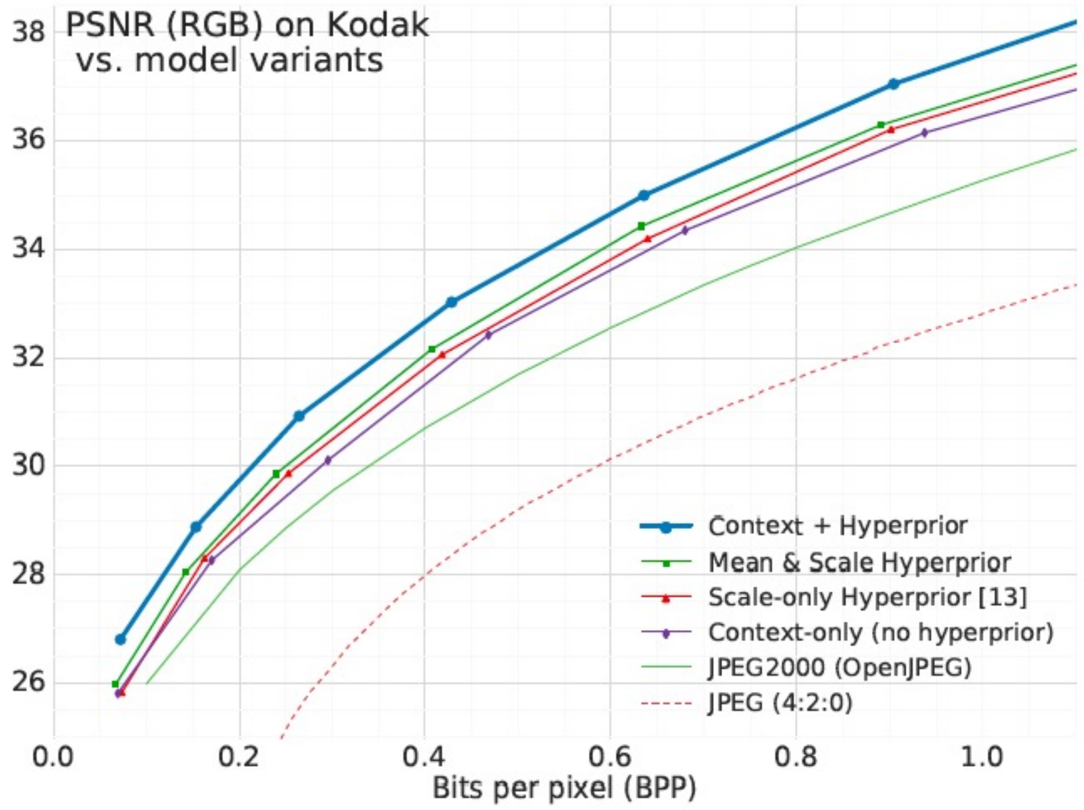
Autoencoder + hyperprior



Autoencoder + hyperprior + context model

Results





Minnen, David, et al. "Joint autoregressive and hierarchical priors for learned image compression", in NeruIPS 2018.

Related Work for Learned Image Compression

[1] Ballé, et al. "End-to-end optimized image compression", ICLR 2017.

[2] Ballé, et al. "Variational image compression with a scale hyperprior", ICLR 2018.

[3] Minnen, et al. "Joint autoregressive and hierarchical priors for learned image compression", NeruIPS 2018.

[4] Toderici, et al. "Variable Rate Image Compression with Recurrent Neural Networks", ICLR 2016.

[5] Toderici, et al. "Full Resolution Image Compression with Recurrent Neural Networks", CVPR 2017.

[6] Johnston, et al. "Improved Lossy Image Compression with Priming and Spatially Adaptive Bit Rates for Recurrent Networks", CVPR 2018.

[7] Agustsson, et al. "Generative adversarial networks for extreme learned image compression", ICCV 2019.

[8] Mentzer, et al. "High-Fidelity Generative Image Compression", NeurIPS 2020.

[9] Choi, et al. "Variable Rate Deep Image Compression With a Conditional Autoencoder", ICCV 2019.

[10] Ma, et al. "End-to-End Optimized Versatile Image Compression With Wavelet-Like Transform", IEEE T-PAMI 2020.

[11] Li et al. "Learning Convolutional Networks for Content-weighted Image Compression", CVPR 2018.

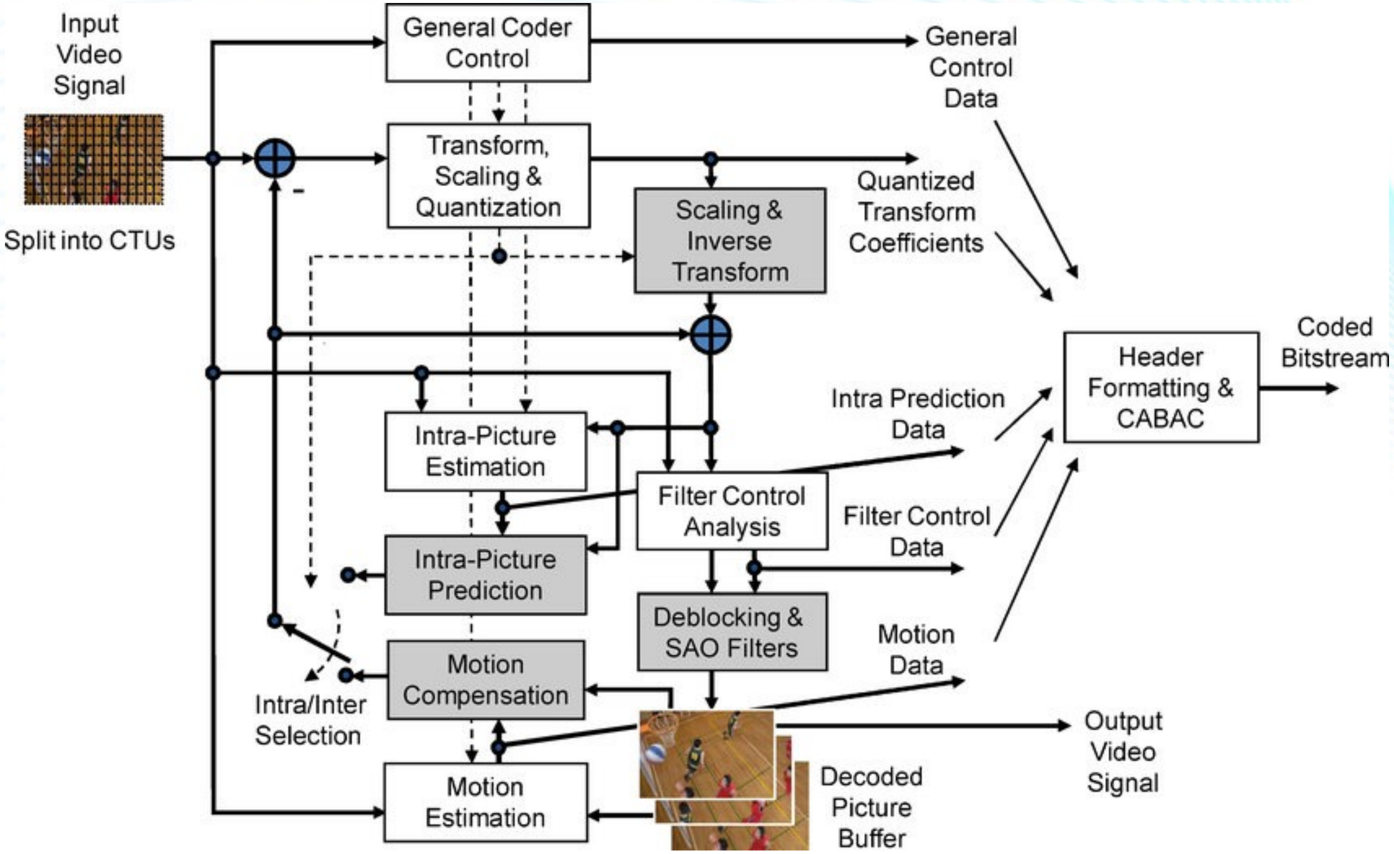
[12] Li et al. "Learning Content-Weighted Deep Image Compression", IEEE PAMI 2020

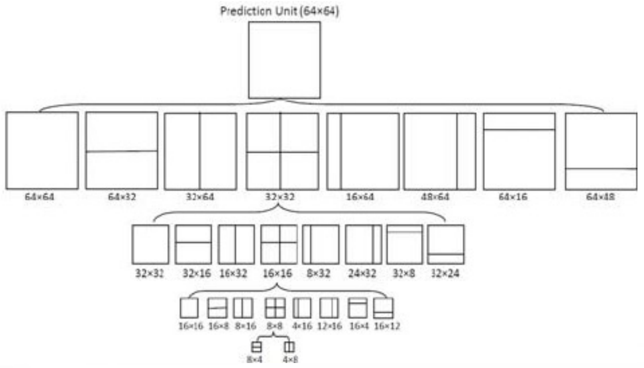
[13] Lee et al. "A Hybrid Layered Image Compressor with Deep-Learning Technique", IEEE MMSP 2020.

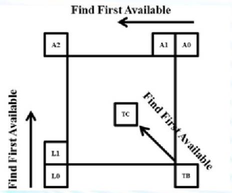
Outline

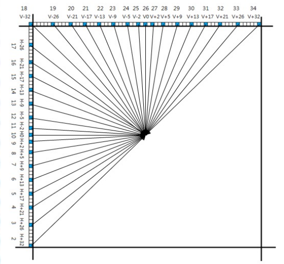
- Overview of learning-based visual data compression
  - Learning-based image compression
  - Learning-based video compression
  - Learning-based volumetric visual data compression
- Standard activities on learning-based visual data compression
  - JPEG AI
  - JVET NNVC
  - IEEE DCSC FVC
- Discussion and related work

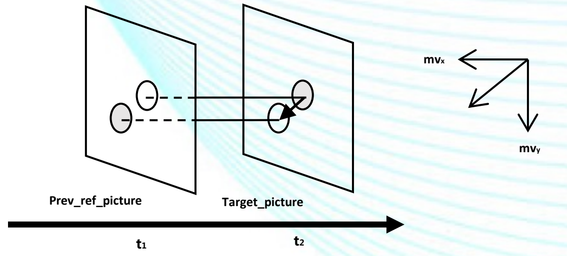
Conventional Video Compression







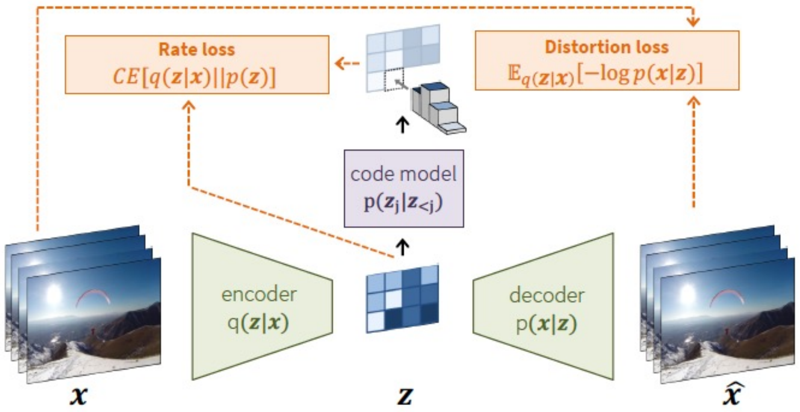




Source: Wikipedia

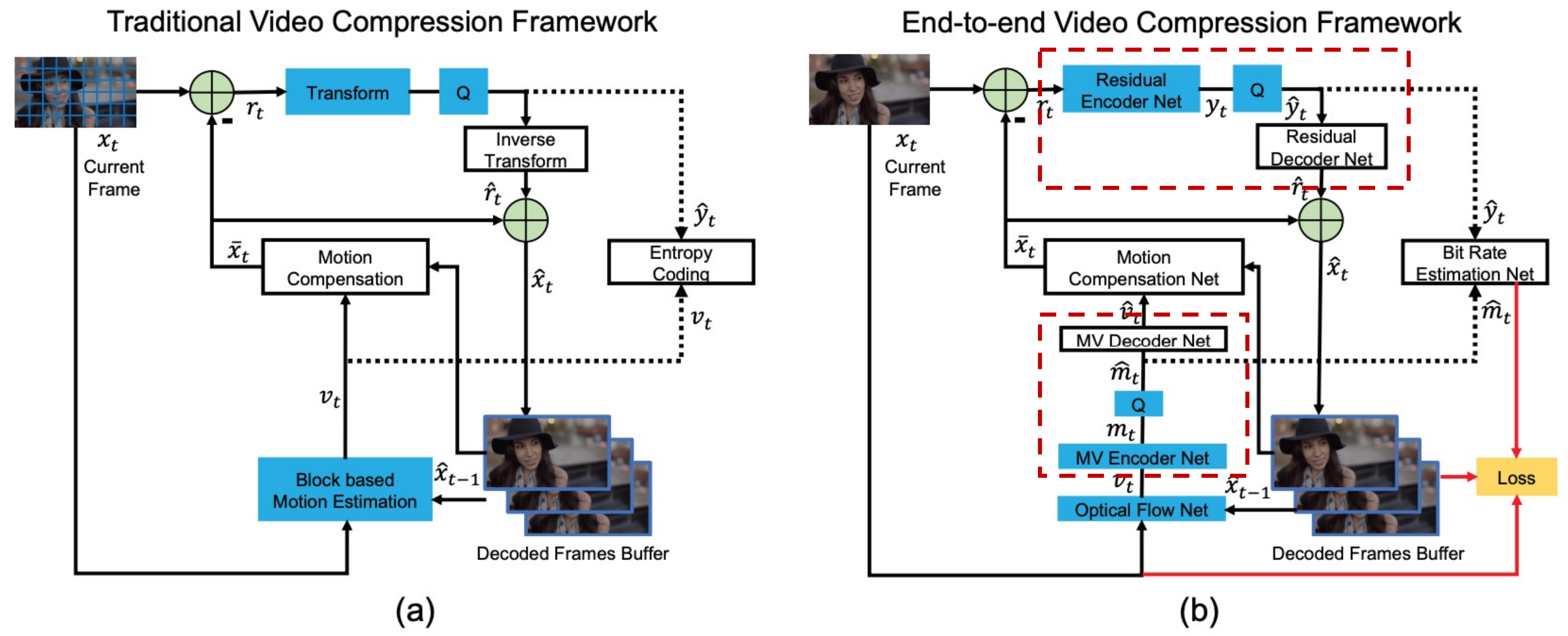
Learning-based Video Compression

- Learned components
  - Learning-based coding tools in conventional compression framework
- Hybrid E2E compression
  - Conventional compression framework with E2E training /joint optimization
- Autoencoder



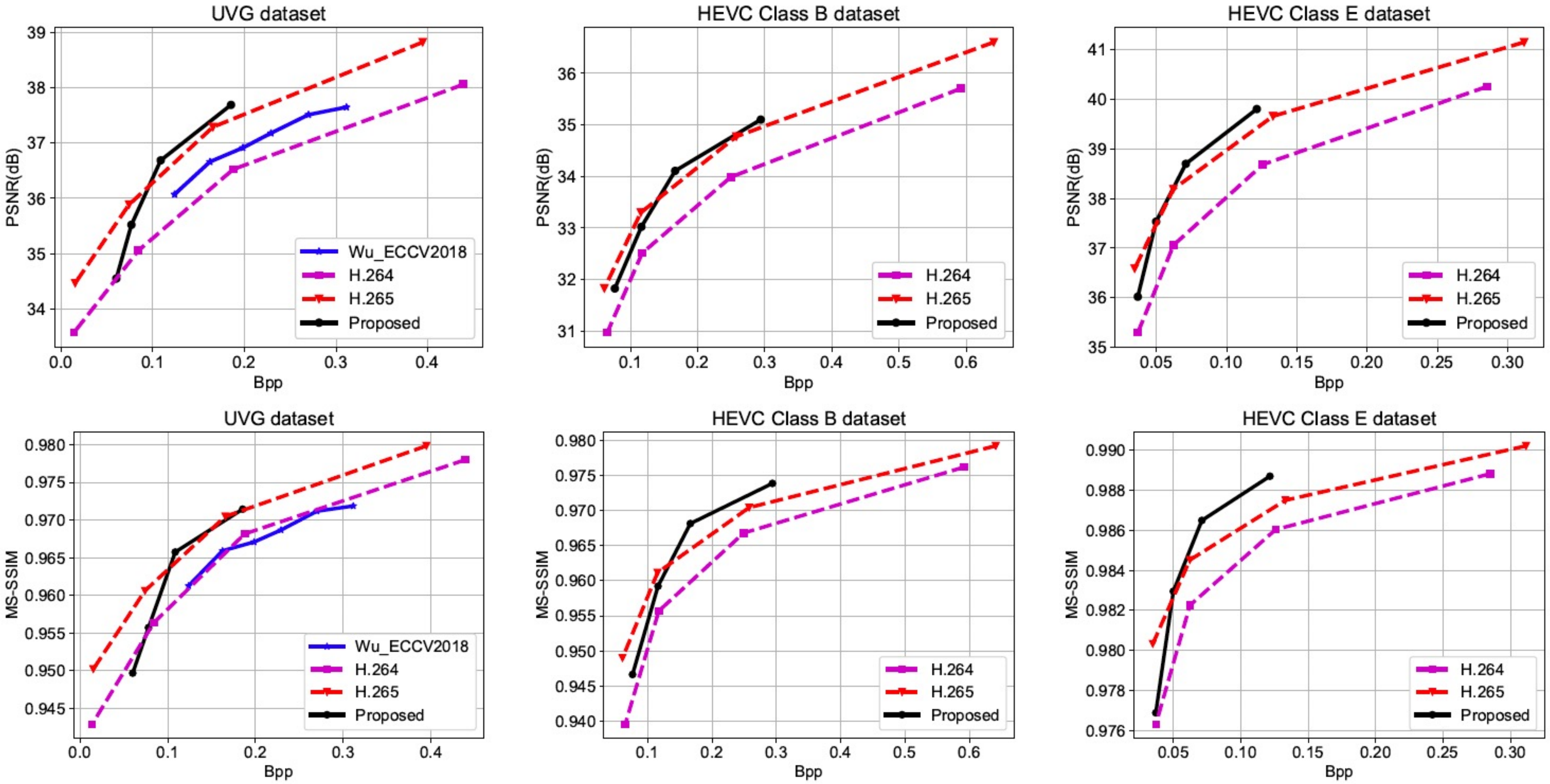
Habibian, et al. "Video compression with rate-distortion autoencoders", ICCV 2019.

E2E Learned Video Compression



Lu et al., “DVC: An end-to-end deep video compression framework,” CVPR 2019.

Results



Lu et al., “DVC: An end-to-end deep video compression framework,” CVPR 2019.

Recent Work on Learned Video Compression

[1] Lu, et al. “DVC: An End-to-end Deep Video Compression Framework”, CVPR 2019.

[2] Lu, et al. “Content Adaptive and Error Propagation Aware Deep Video Compression”, ECCV 2020.

[3] Hu, et al. “Improving Deep Video Compression by Resolution-adaptive Flow Coding”, ECCV 2020.

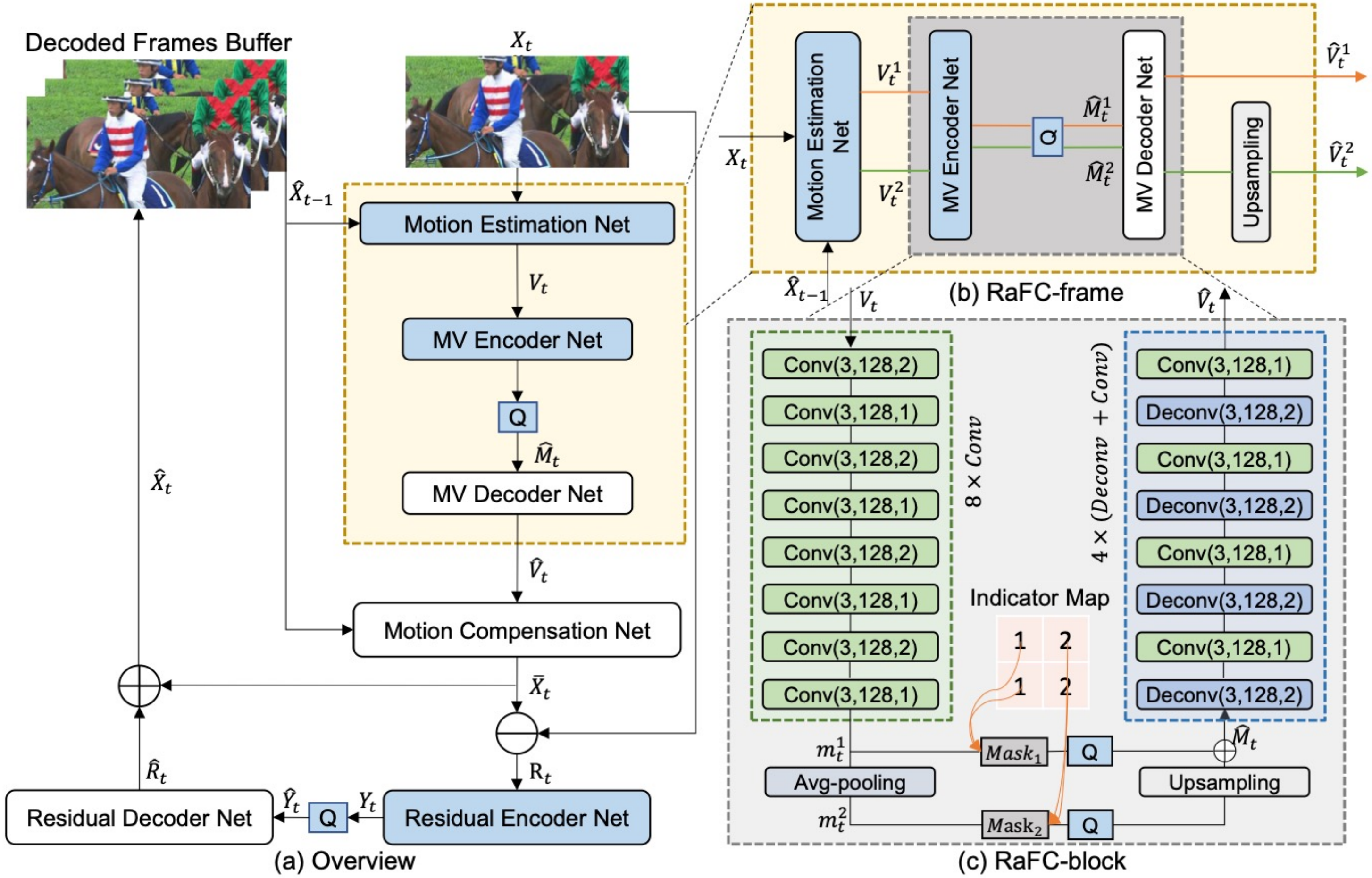
[4] Agustsson, et al. "Scale-Space Flow for End-to-End Optimized Video Compression”, CVPR 2020.

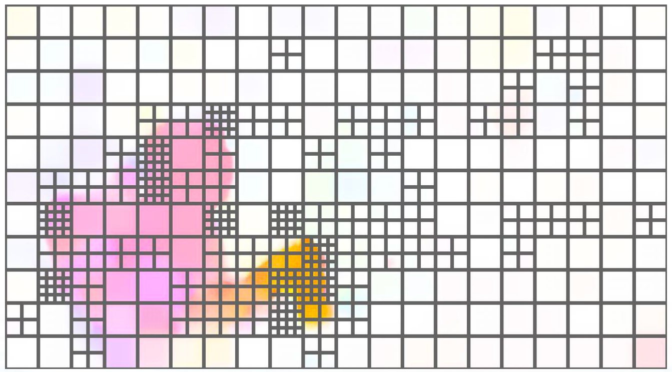
[5] Lin, et al. "M-LVC: Multiple Frames Prediction for Learned Video Compression”, CVPR 2020.

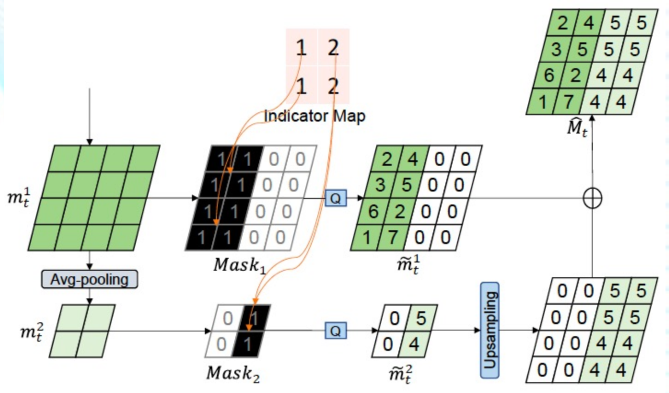
[6] Yang, et al. "Learning for video compression with hierarchical quality and recurrent enhancement”, CVPR 2020.

[7] Habibian, et al. "Video compression with rate-distortion autoencoders”, ICCV 2019.

RaFC

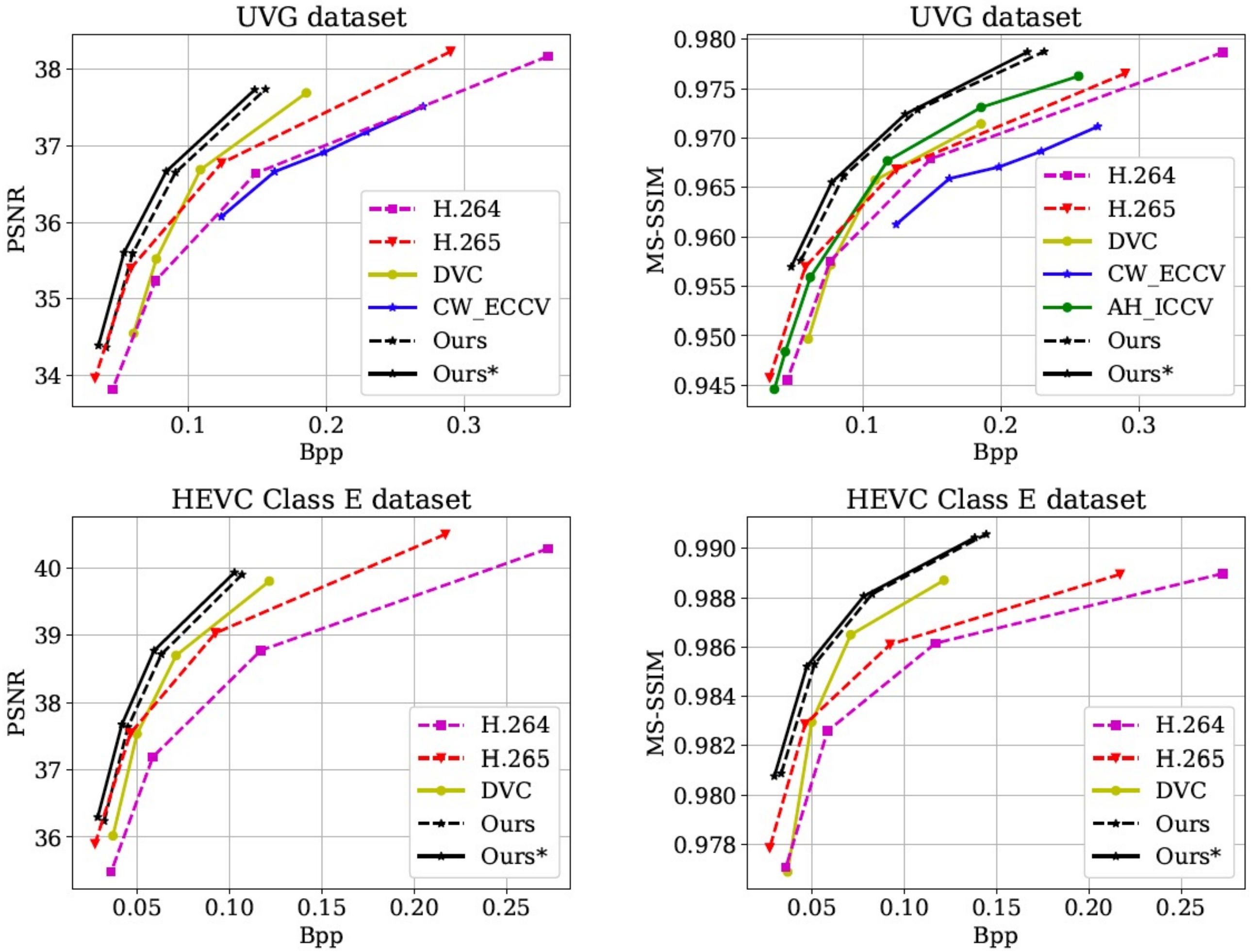






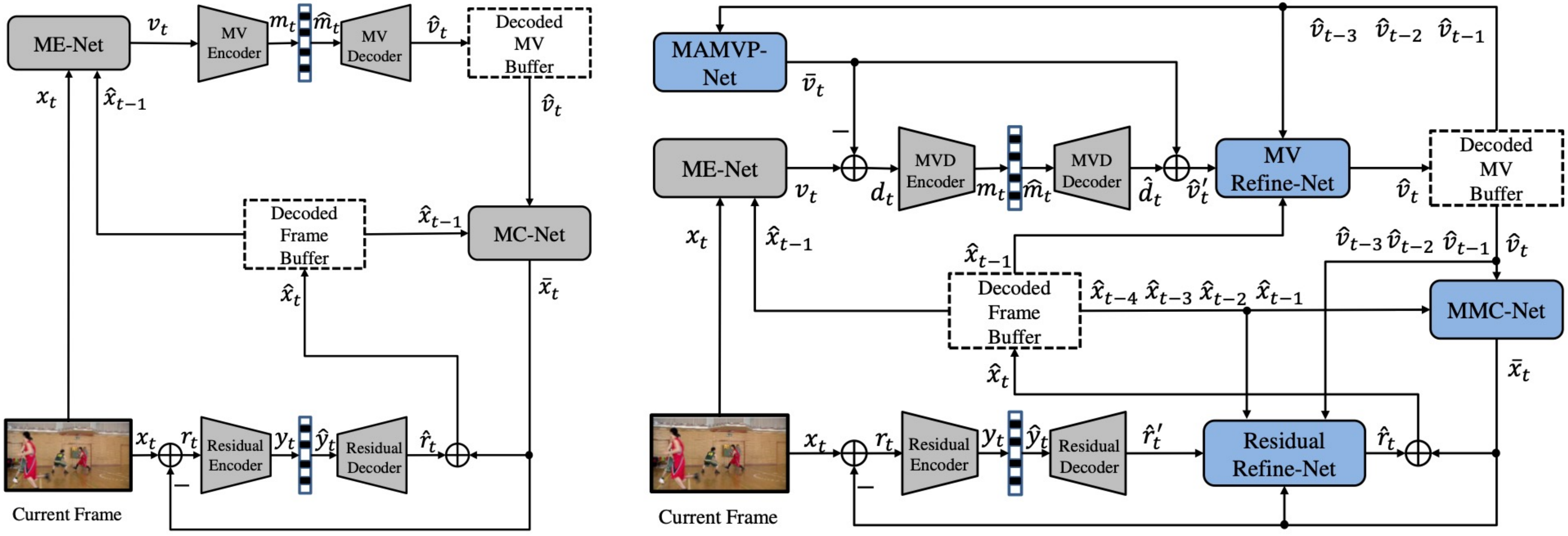
Hu, et al. “Improving Deep Video Compression by Resolution-adaptive Flow Coding”, ECCV 2020.

RaFC Results



Hu, et al. "Improving Deep Video Compression by Resolution-adaptive Flow Coding", ECCV 2020.

M-LVC

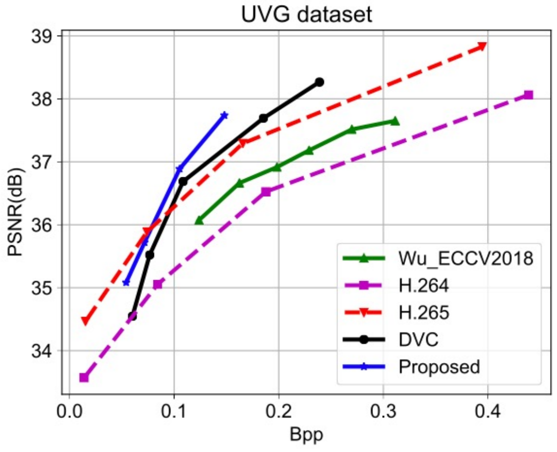


(a) DVC

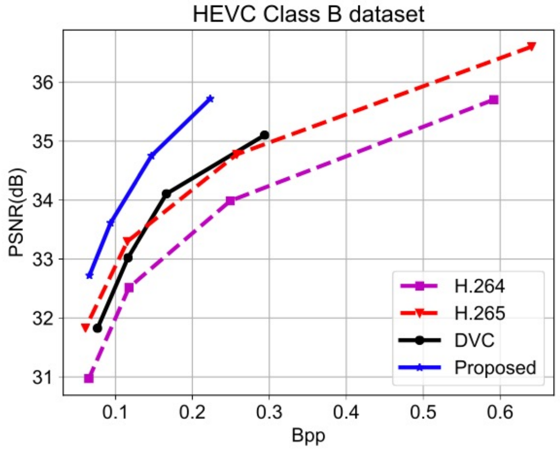
(b) M-LVC

Lin, et al. "M-LVC: Multiple Frames Prediction for Learned Video Compression", CVPR 2020.

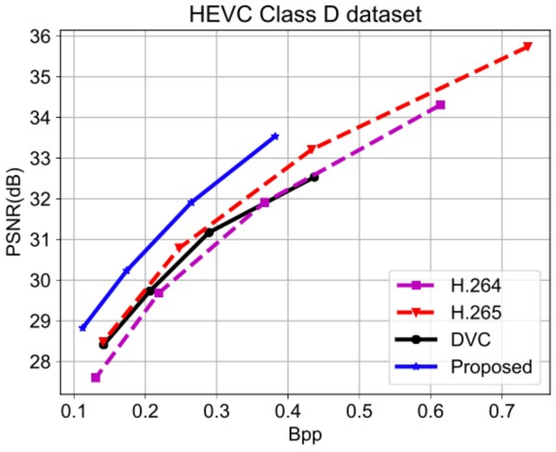
M-LVC Results



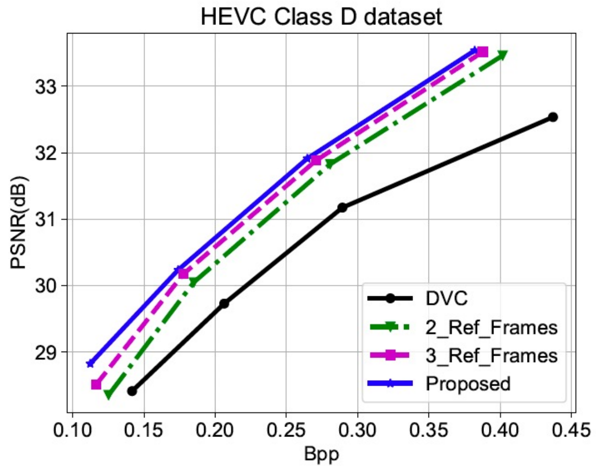
(a)

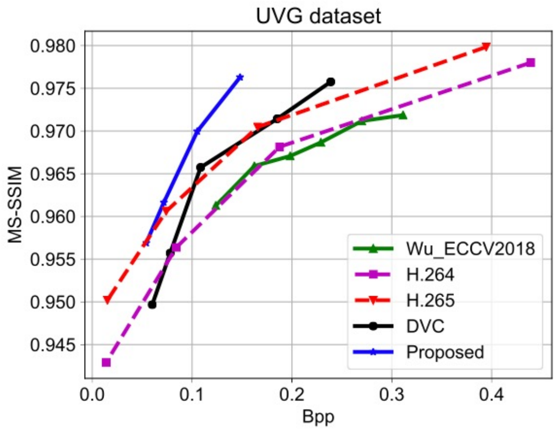


(b)

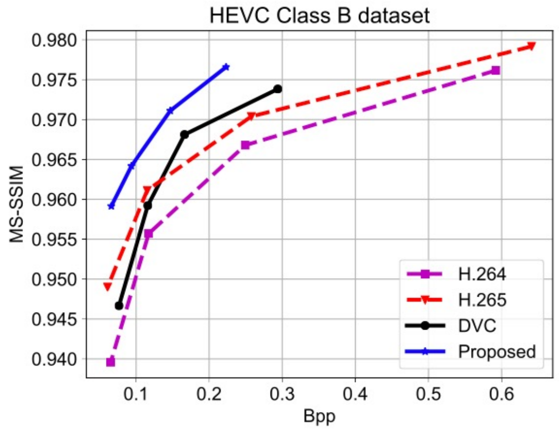


(c)

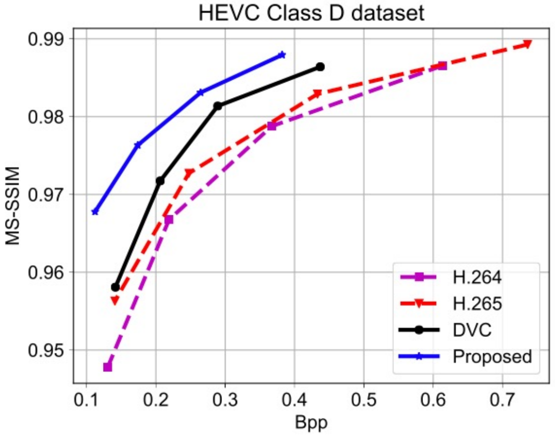




(d)



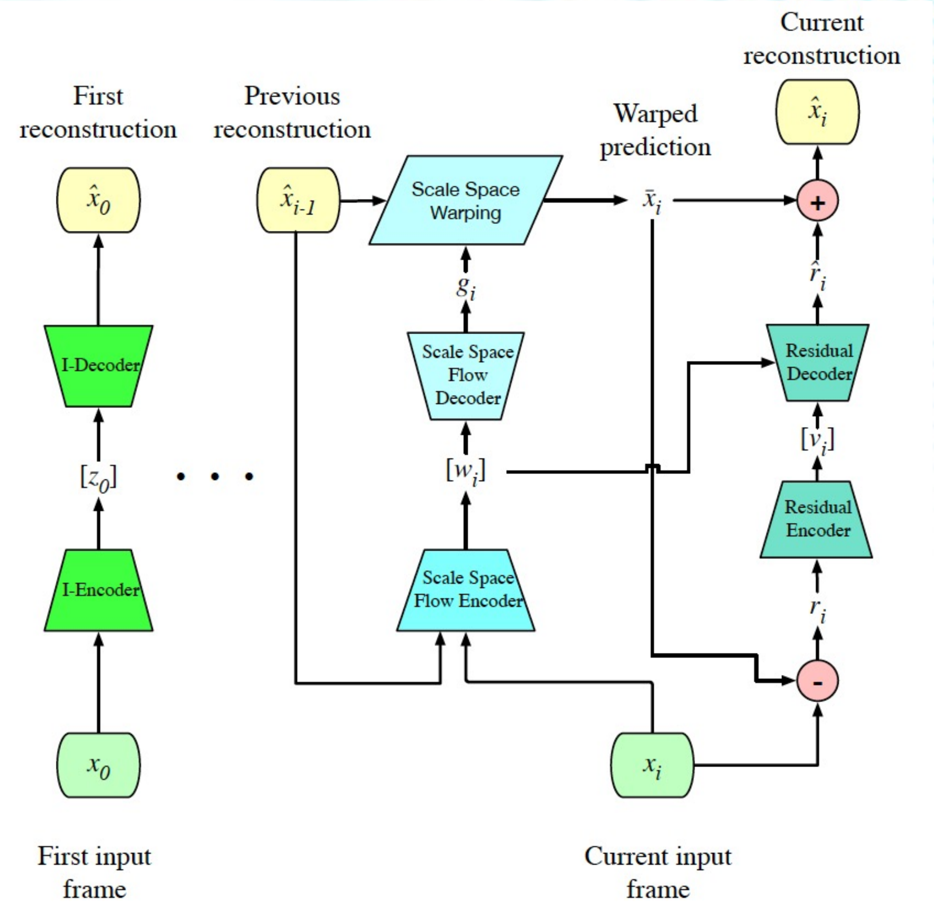
(e)

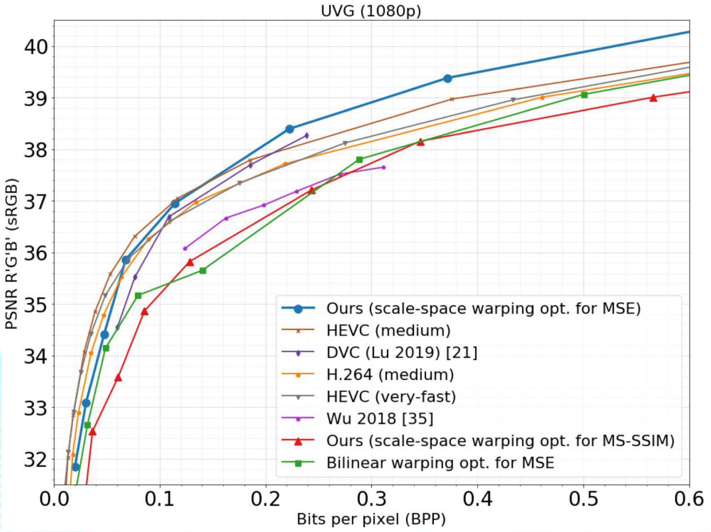


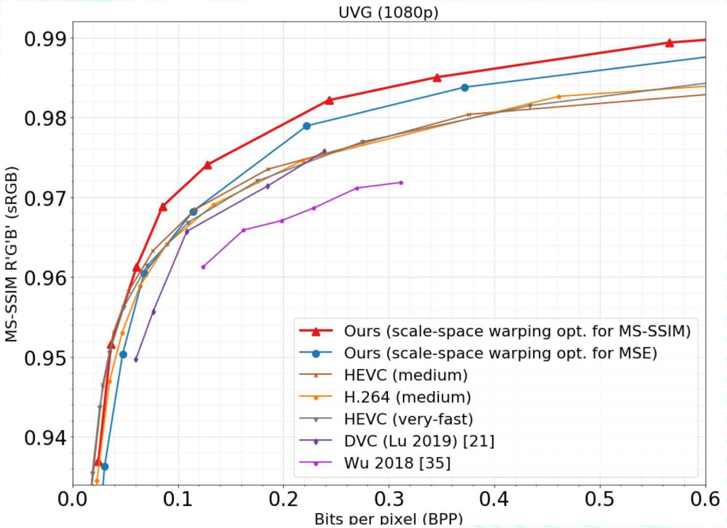
(f)

Lin, et al. "M-LVC: Multiple Frames Prediction for Learned Video Compression", CVPR 2020.

Scale-Space Flow

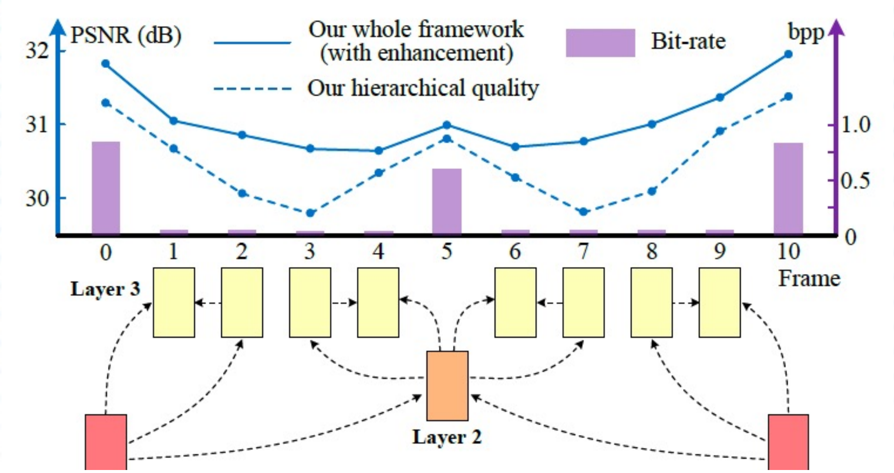


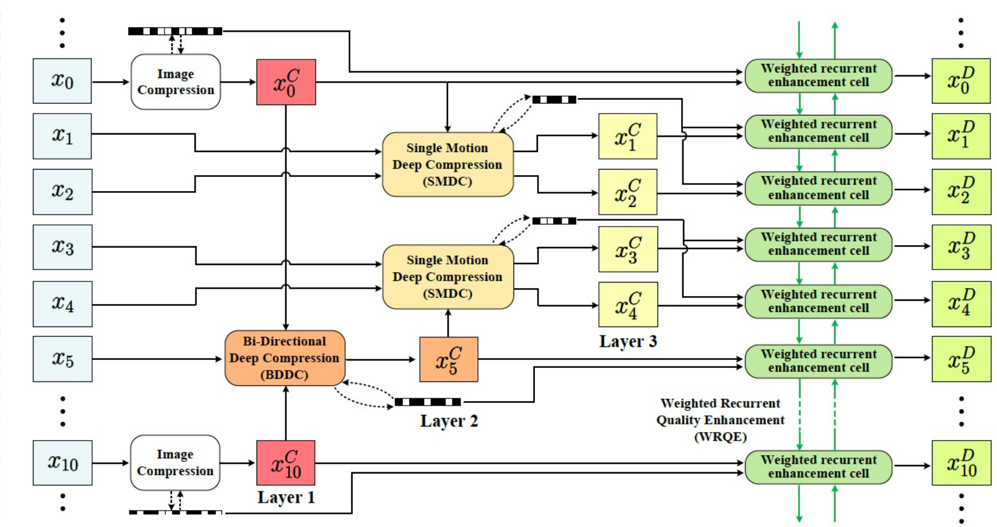


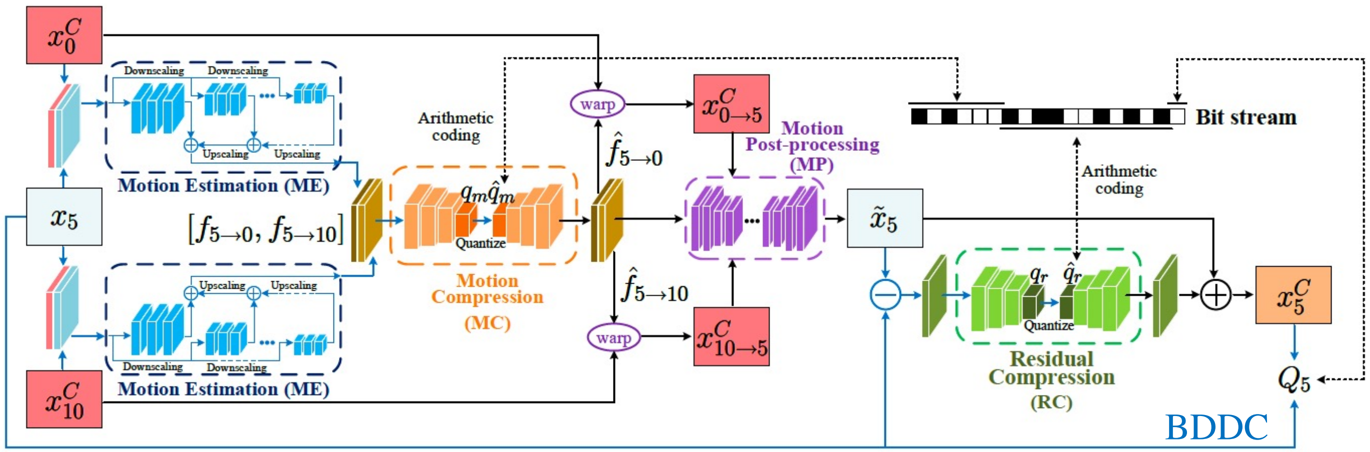


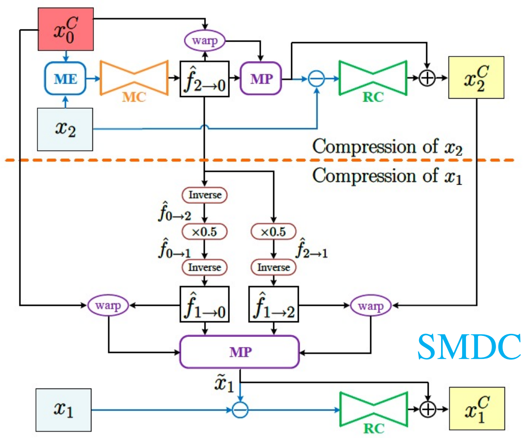
Agustsson, et al. "Scale-Space Flow for End-to-End Optimized Video Compression", CVPR 2020.

HLVC



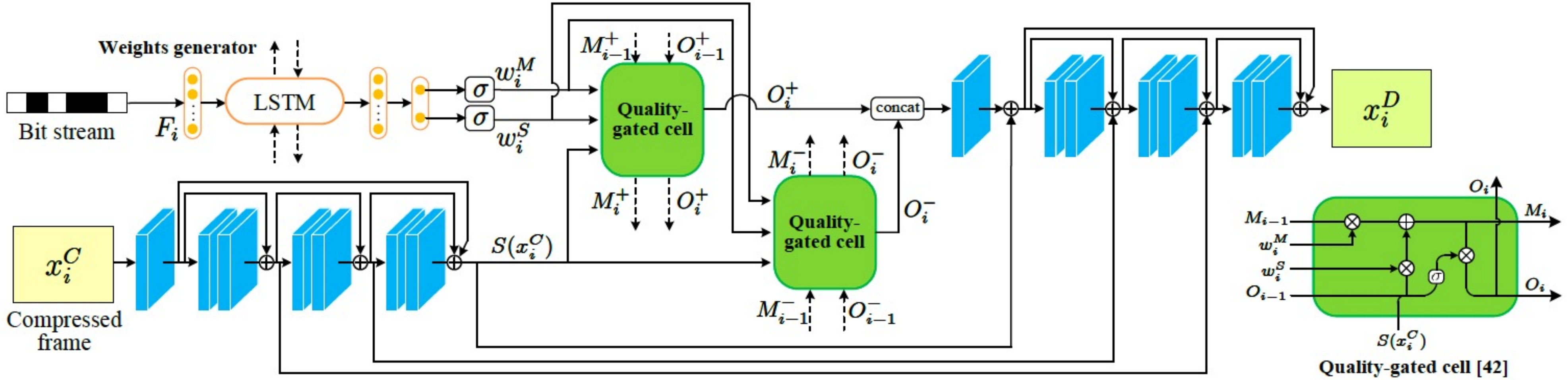




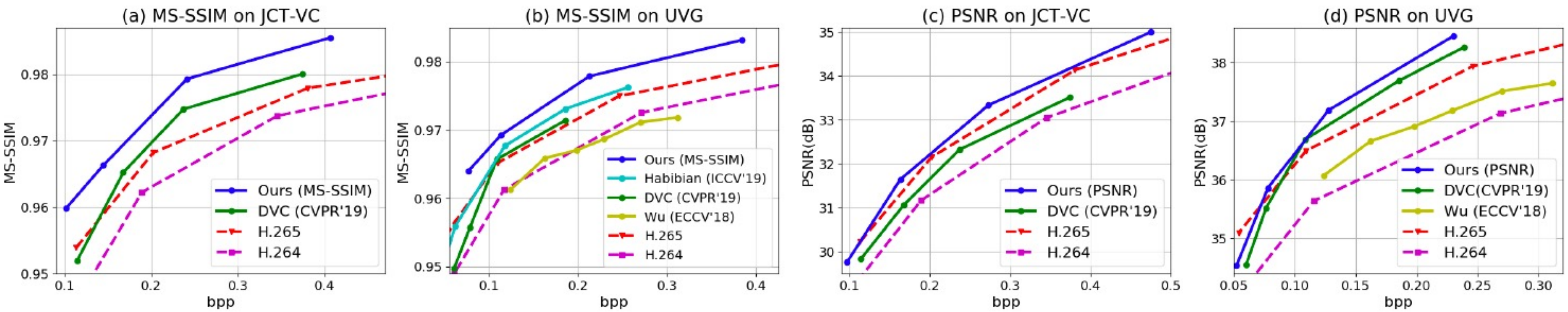


Yang, et al. "Learning for video compression with hierarchical quality and recurrent enhancement", CVPR 2020.

HLVC

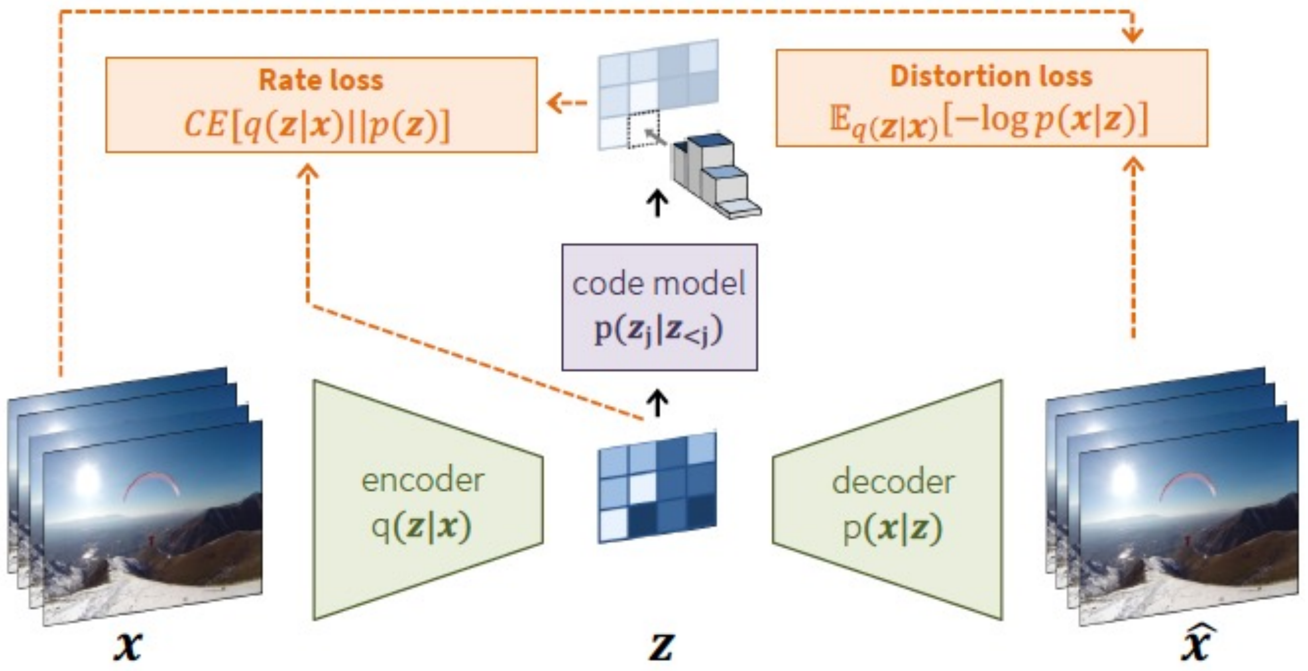


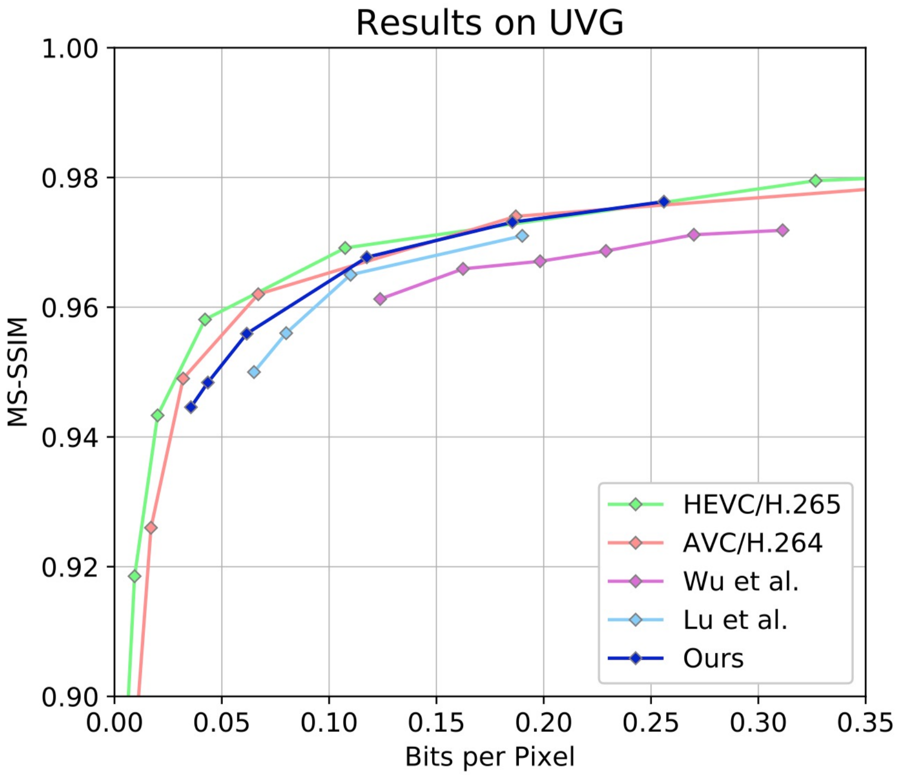
WRQE



Yang, et al. "Learning for video compression with hierarchical quality and recurrent enhancement", CVPR 2020.

Autoencoder





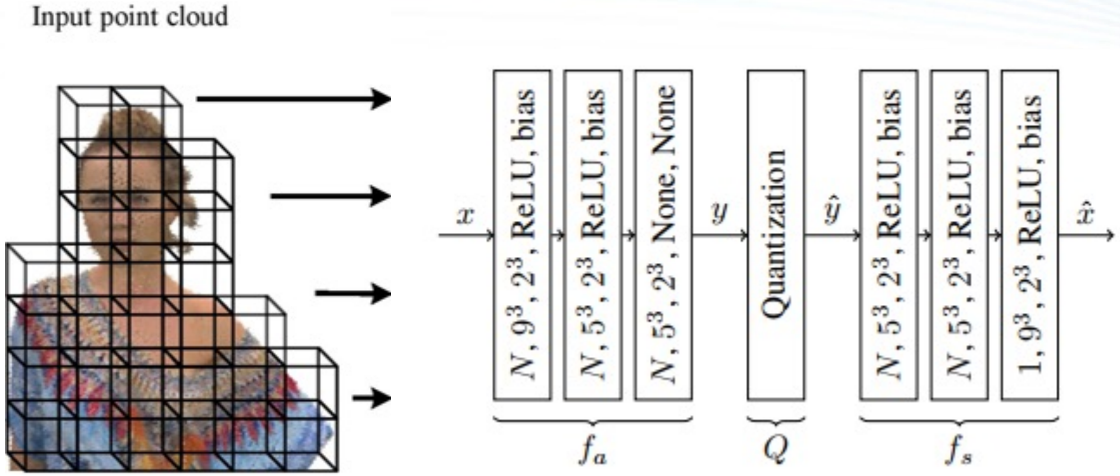
Habibian, et al. "Video compression with rate-distortion autoencoders", ICCV 2019.

Outline

- Overview of learning-based visual data compression
  - Learning-based image compression
  - Learning-based video compression
  - Learning-based volumetric visual data compression
- Standard activities on learning-based visual data compression
  - JPEG AI
  - JVET NNVC
  - IEEE DCSC FVC
- Discussion and related work

Learned Point Cloud Compression

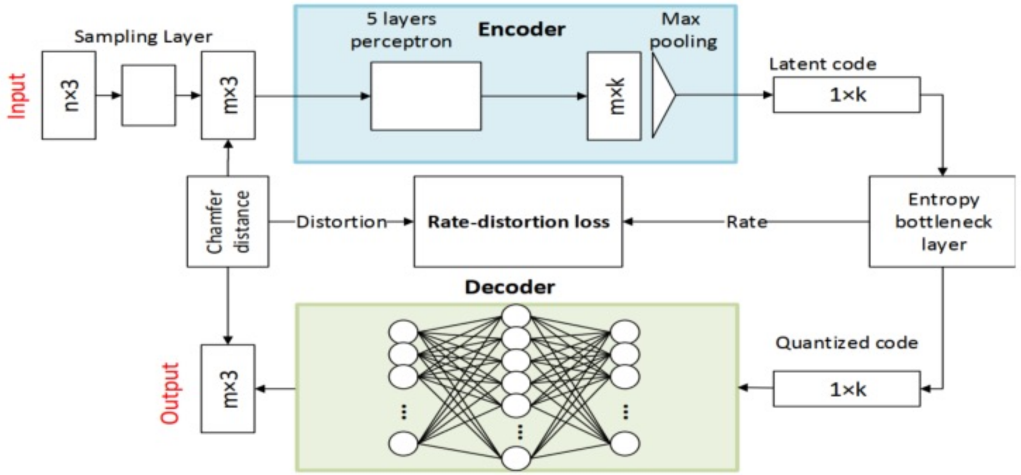
- Motivated by the framework used in learning-based image compression
  - Auto-encoder with an analysis network f_a and a synthesis network f_s
  - Extracted feature y are quantized into $\hat{y}$ and are encoded
  - 2D convolution/deconvolution ➔ 3D dense convolution/deconvolution
  - 2D images ➔ 3D voxelized space
    - Which is partitioned into multiple non-overlapping cubes that are fed into the network



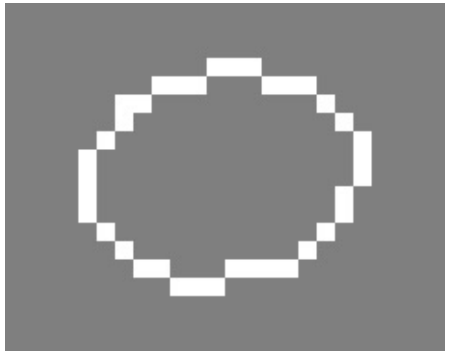
M. Quach, et al. "Learning convolutional transform for lossy point cloud geometry compression", ICIP2019.

Learned Point Cloud Compression

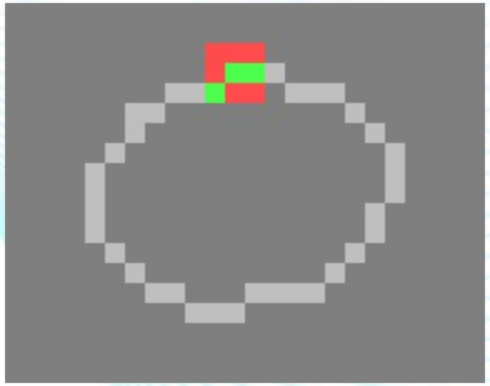
- Voxel representation and dense convolution
  - Incur high computational complexity and large memory usage
  - Can only partition a point cloud into small blocks for processing, thus limit performance improvement
- Improved schemes were proposed for lower complexity and better performance
  - Point-based processing
  - (Submanifold) sparse convolution



W. Yan, et al. "Deep AutoEncoder-based Lossy Geometry Compression for Point Clouds", arXiv 2019.



With regular 3x3 convolutions, the set of active (non-zero) sites grows rapidly



With submanifold sparse convolution, the set of active sites is unchanged. Active sites look at their active neighbors (green); non-active sites (red) are not used.

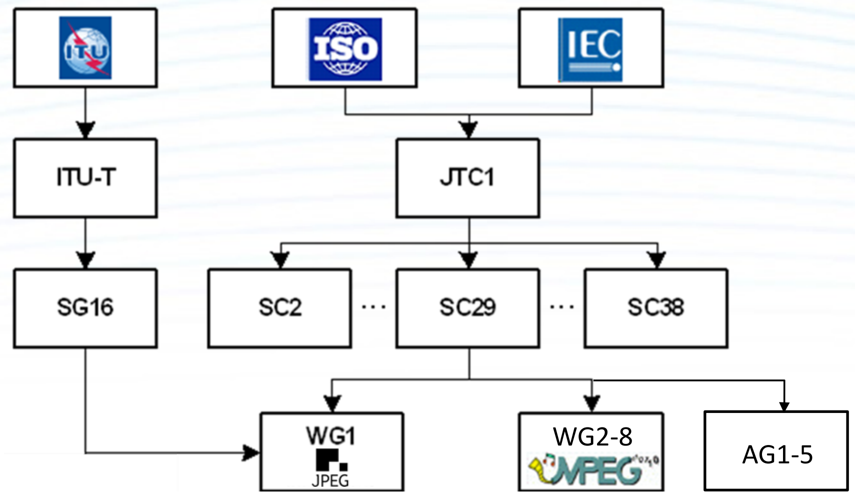
https://github.com/facebookresearch/SparseConvNet

Outline

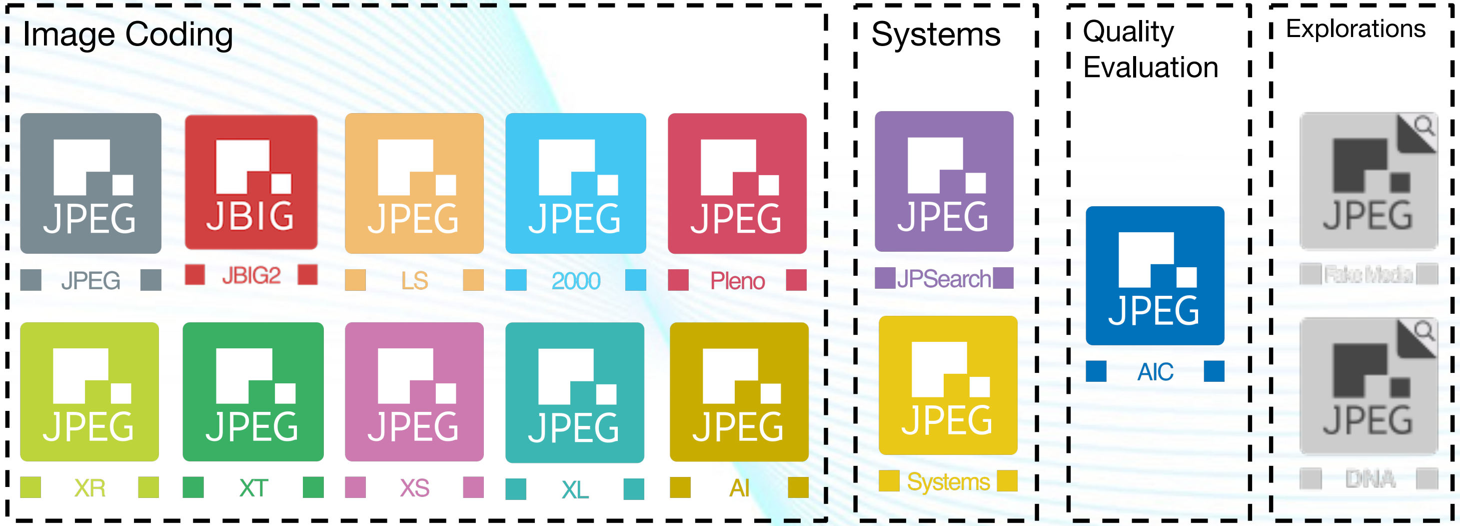
- Overview of learning-based visual data compression
  - Learning-based image compression
  - Learning-based video compression
  - Learning-based volumetric visual data compression
- Standard activities on learning-based visual data compression
  - JPEG AI
  - JVET NNVC
  - IEEE DCSC FVC
- Discussion and related work

Introduction to JPEG AI

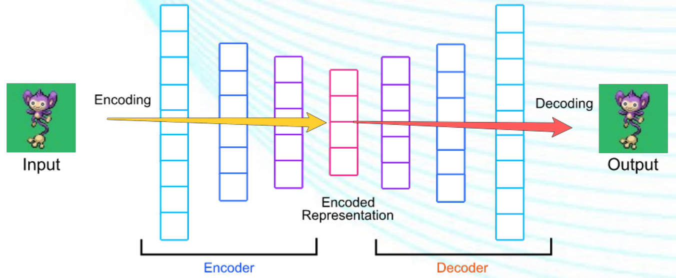
- What is JPEG?



JPEG family standards



- JPEG AI
  - Auto-encoder (as opposed to learning-based tools or components in conventional coding architecture)

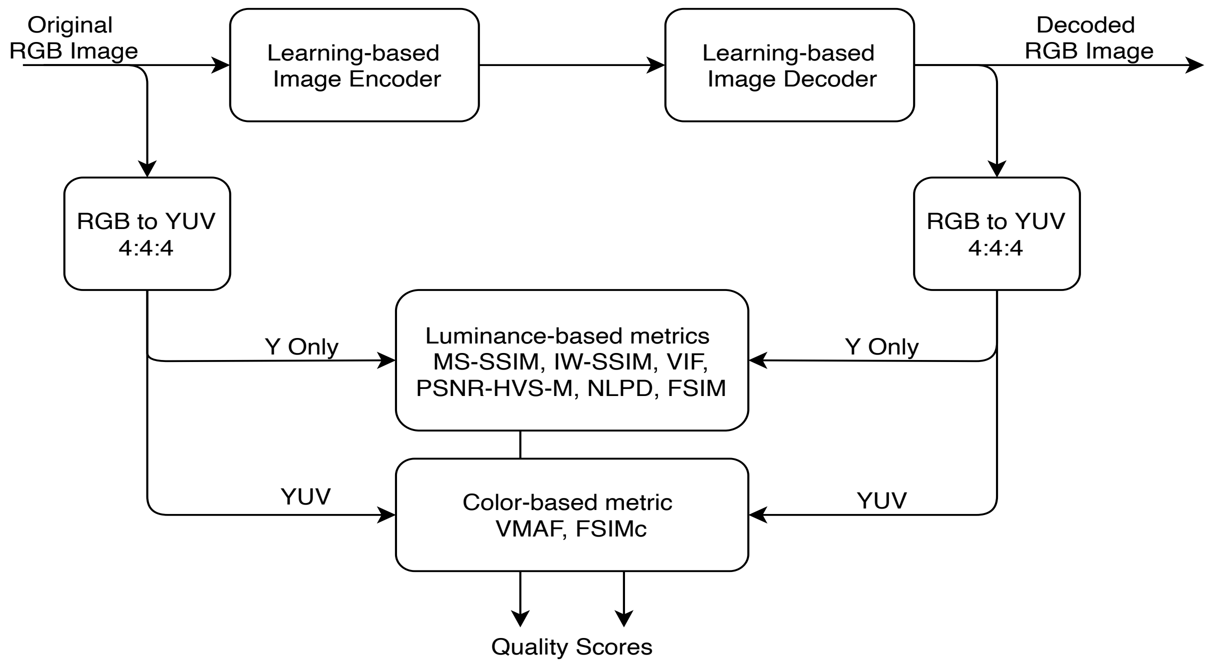


JPEG AI History and Timeline

- **January 2019**: Establishment of an AHG on learning-based image coding as an Exploration activity
- **March 2019**: A first public report on state of the art in learning-based image coding
- **November 2019**: A first complete objective and subjective assessment of the state-of-the-art learning-based image coding
- **February 2020**: A Call for Evidence issued combined with the IEEE MMSP Workshop Grand Challenge
  - 6 codecs submitted (out of 8 registered)
- **October 2020**: Final report of the call for Evidence and decision to initiate JPEG AI as a New Work Item
- **January 2021**: A draft Call for Proposal.
- **April 2021**: Call for Proposal

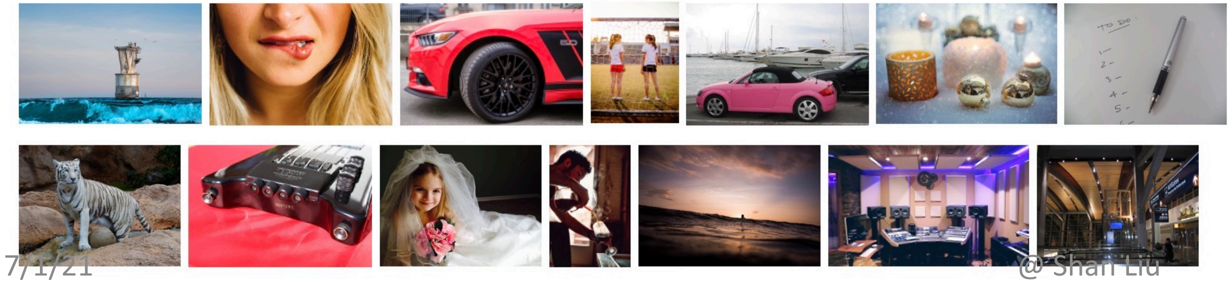
JPEG AI CfE Test Conditions

- Target rates (JPEG N90055)
  - bpp @ {0.06, 0.12, 0.25, 0.50, 0.75, 1.00, 1.50, and 2.00}
  - Max deviation <=15%
- Evaluation Procedure
  - Objective metric: MS-SSIM, VMAF, VIFP, NLPD, FSIM
  - Subjective evaluation:
    - Critical since the type of artifacts that learning-based image compression may be different from standard image codecs.
    - Double Stimulus Continuous Quality Scale with 5-point scale
    - Four bitrate points covering a wide range of qualities will be used

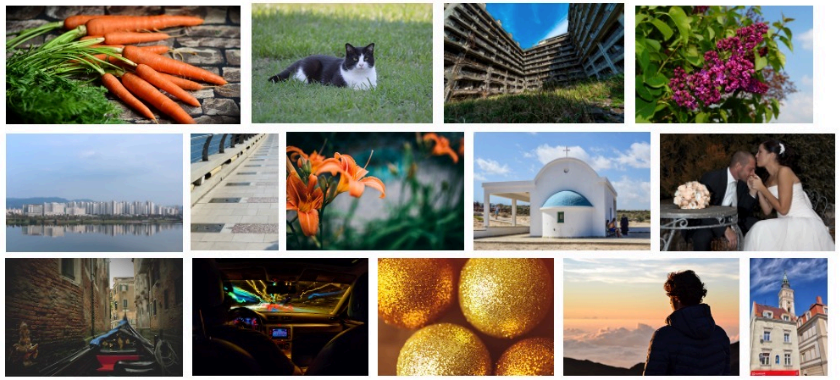


JPEG AI CfE Dataset

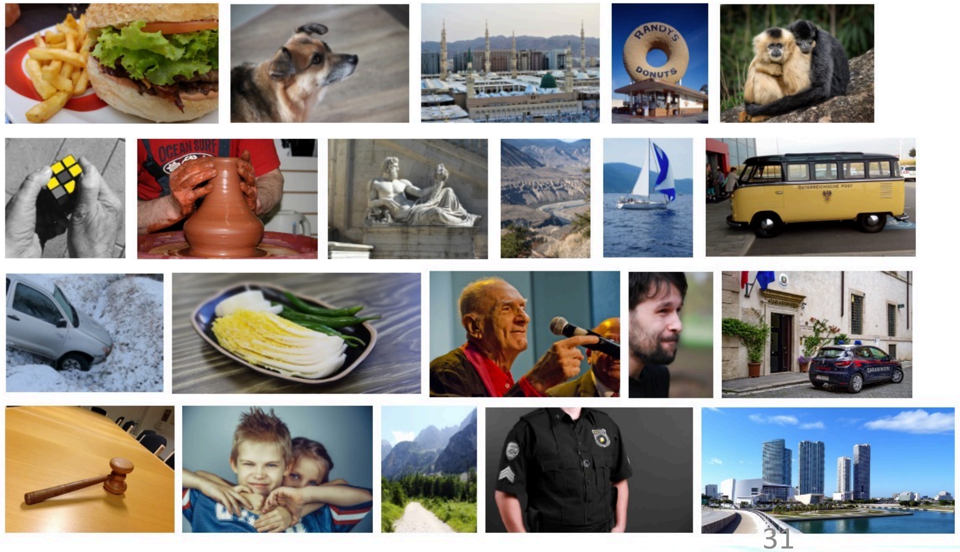
- Training/validation dataset (right)
  - PNG images (RGB)
  - 256×256 to 8K (8 bit) resolution
  - 5264/350 images (righ thand side)
- Test dataset (hidden during CfE)
  - PNG images (RGB)
  - 960x642 to 6016x4016 (8 bit) resolution
  - 40 images (below)



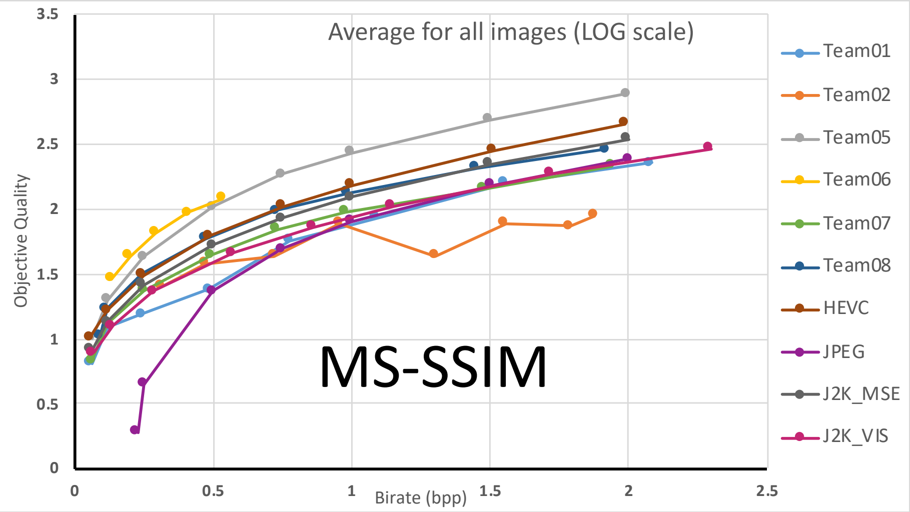
Validation set

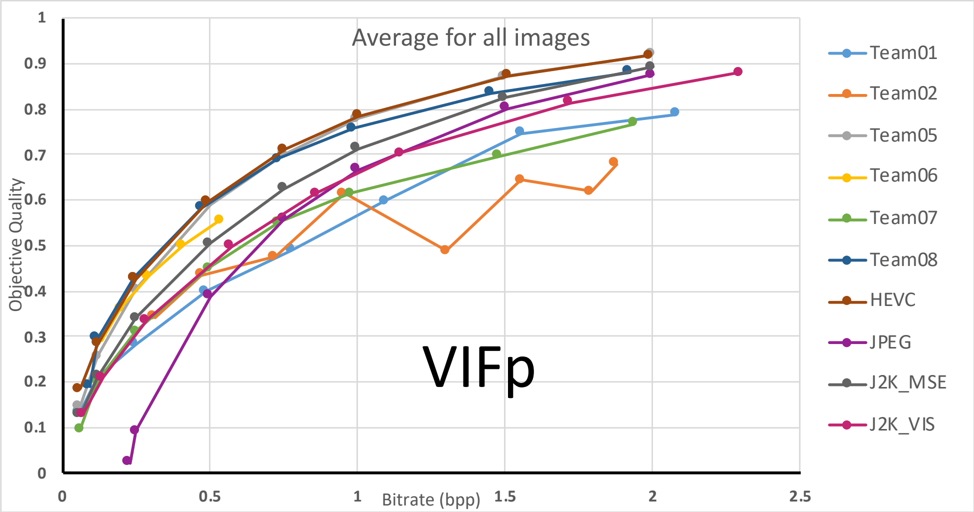


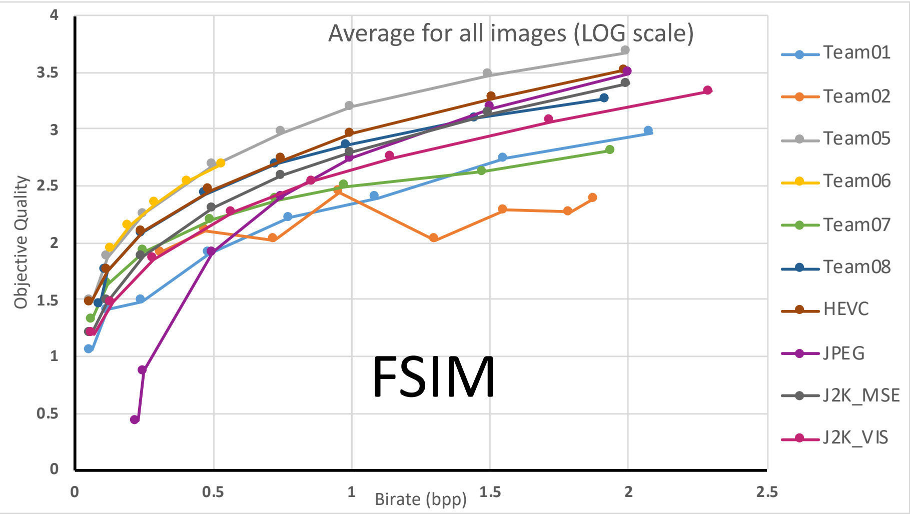
Training set

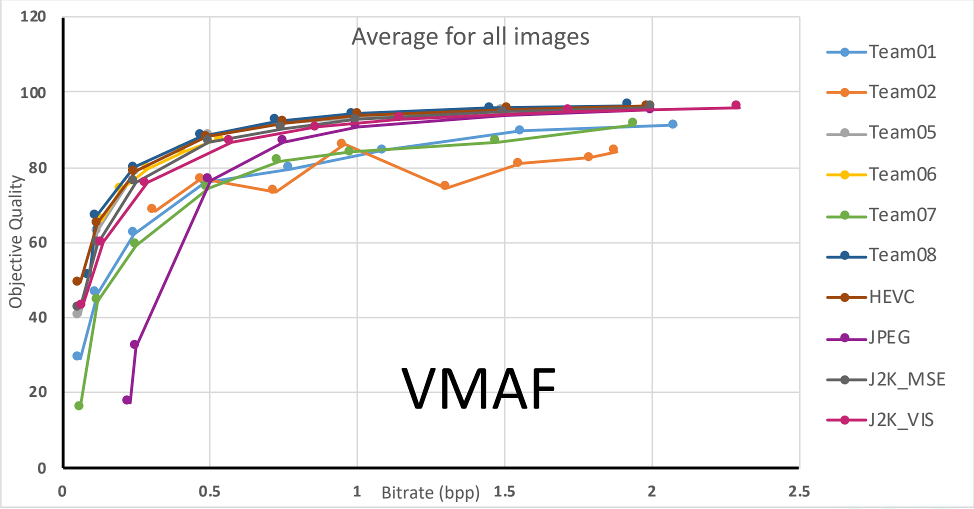


JPEG AI CfE Results (objective metrics)

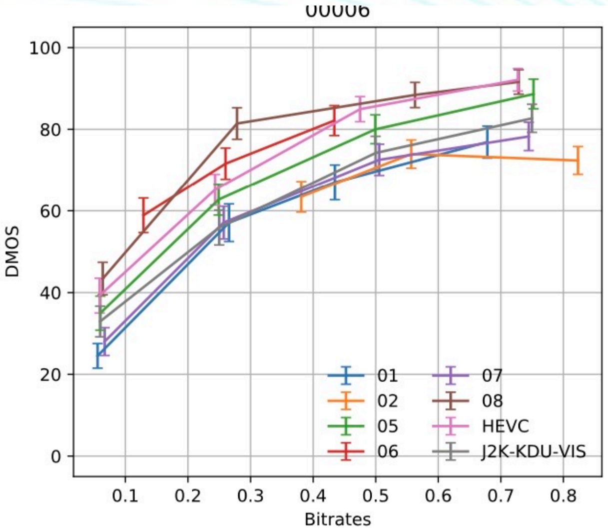


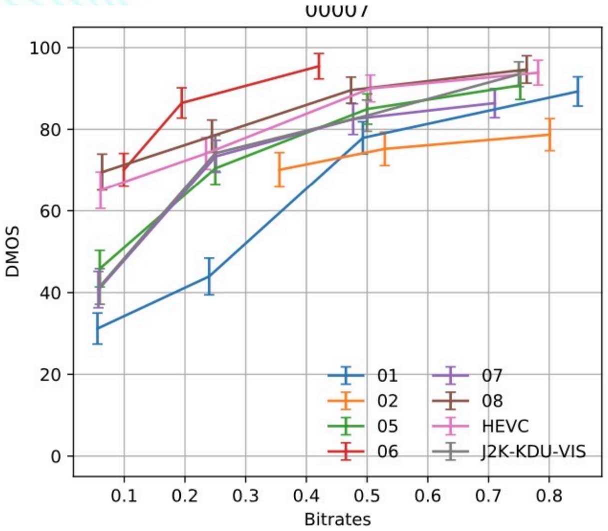


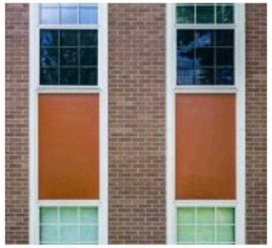


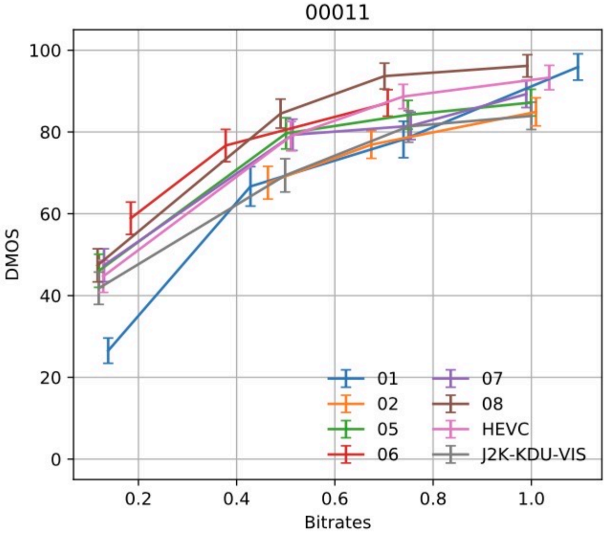


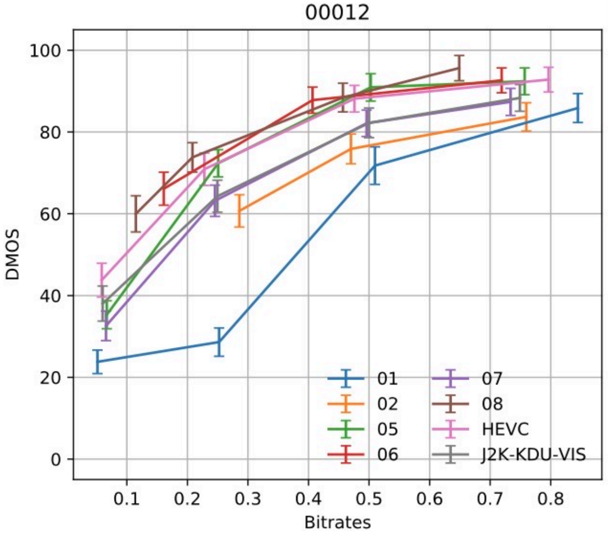
JPEG AI CfE Results (subjective metrics)

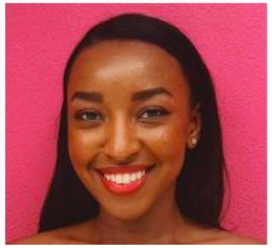








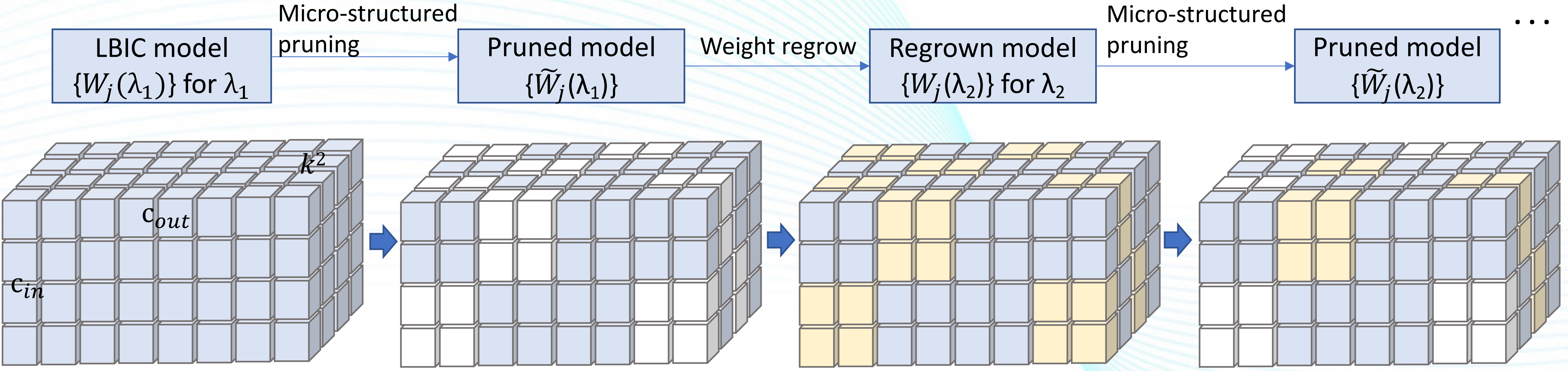




JPEG AI AhG

Subgroup	Requirements
Status	Public
Name	**Ad Hoc Group on JPEG AI**
Chair	João Ascenso (joao.ascenso@lx.it.pt)
Co-chair	Evgeniy Upenik (evgeniy.upenik@epfl.ch)
Mandates/Objectives	•Perform exploration study on compressed domain computer vision and image processing tasks •Improve the JPEG AI Use cases and Requirements •Improve the JPEG AI Common Test Conditions •Improve the draft JPEG AI Call for Proposals
Deliverables	•Proposed Draft JPEG AI Call for Proposals •Proposed revision of JPEG AI Use cases and Requirements •Proposed revision of JPEG AI Common Test Conditions •Report on the exploration study on compressed domain computer vision and image processing tasks
Meetings	•Online meeting on 2021-02-17, 13:00 UTC •Online meeting on 2021-03-17, 13:00 UTC •Online meeting on 2021-04-14, 13:00 UTC Additional online meetings may be scheduled as required with at least one week's notice.
How to join	E-mail reflector: jpeg-ai To subscribe to the reflector, please visit http://listregistration.jpeg.org or in case of problems contact lists@jpeg.org.

JPEG AI (variable bit-rate coding)



Mask $\{M_{1j}\}$ **for** λ_1 $\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} \iddots$

Mask $\{M_{2j}\}$ **for** λ_2 $\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} \iddots$

W. Jiang, et al. “Multi-Rate Learning-Based Image Coding with Micro-Structured Masks”, ISO|IEC MPEG wg1, m89074, Oct. 2020.
W. Jiang, et al. “PnG: Micro-structured Prune-and-Grow Networks for Flexible Image Restoration”, NTIRE 2021.

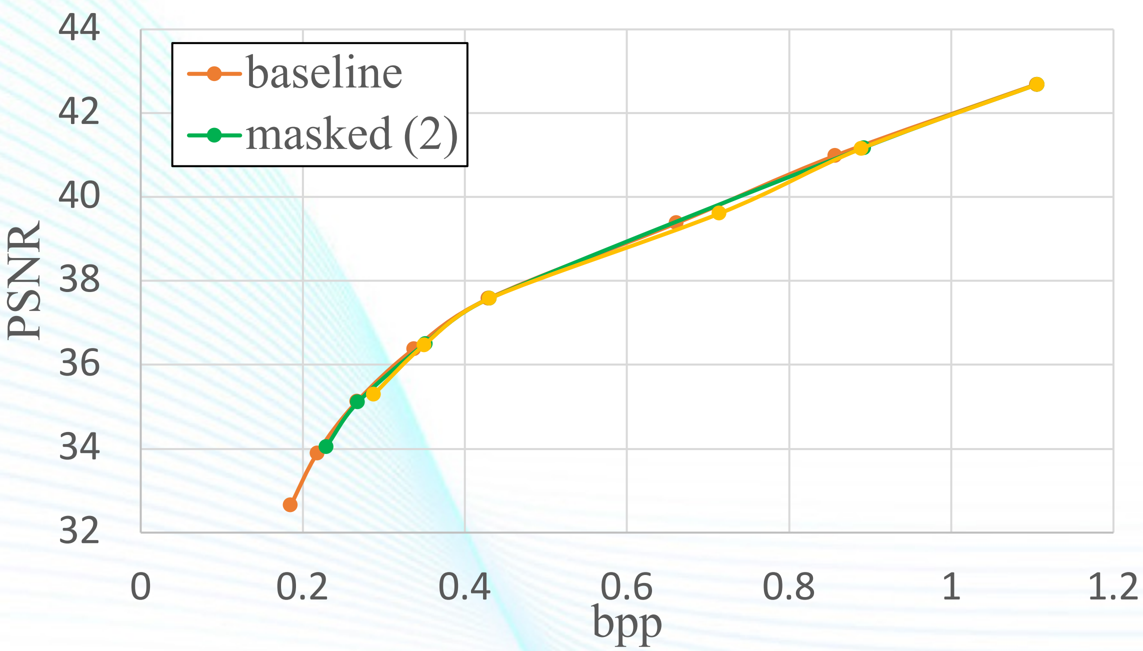
Results

2-task (bitrate) model

3-task (bitrate) model

Less than 1% performance drop

About 50% model reduction



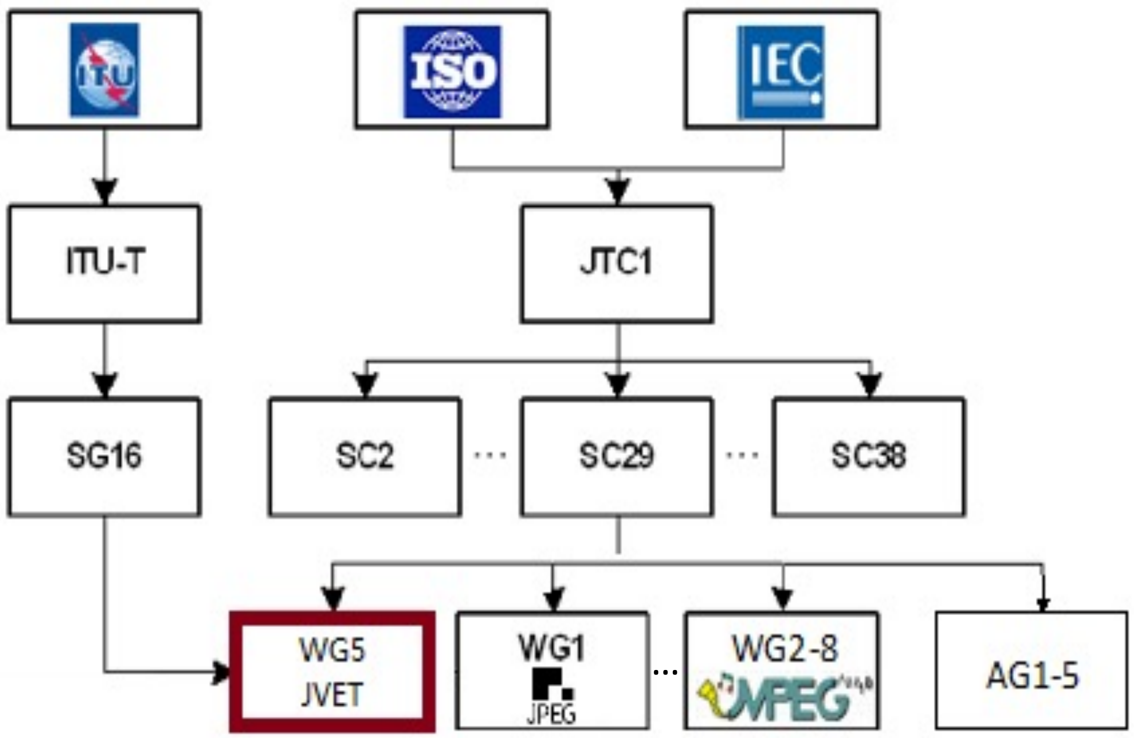
Statistics of model parameters

	# shared			# independent			# used		
λ	Baseline	masked (2)	masked (3)	baseline	masked (2)	masked (3)	baseline	masked (2)	masked (3)
0.18	0	10076160	10076160	11,816,323	1740163	1740163	11,816,323	11190151	1190151
0.0932	0	10076160	10076160	11,816,323	1740163	1740163	11,816,323	11,816,323	11644500
0.0483	0	10076160	10076160	11,816,323	1740163	1740163	11,816,323	--	11,816,323
0.025	0	4317184	4317184	5,075,843	758659	758659	5,075,843	4803991	4803991
0.013	0	4317184	4317184	5,075,843	758659	758659	5,075,843	5,075,843	504892
0.0067	0	4317184	4317184	5,075,843	758659	758659	5,075,843	4807175	5,075,843
0.0035	0	4317184	4317184	5,075,843	758659	758659	5,075,843	5,075,843	--
0.0018	0	4317184	4317184	5,075,843	758659	758659	5,075,843	--	--

W. Jiang, et al. “Multi-Rate Learning-Based Image Coding with Micro-Structured Masks”, ISO|IEC MPEG WG1, m89074, Oct. 2020.

Introduction to JVET NNVC

- What is JVET?
  - Joint Video Expert Team by ITU-T SG16 VQEG and ISO/IEC JTC 1/SG 29 MPEG WG5
  - Standard committee that has been responsible for H.266/VVC standardization
- ITU-T/SG 16 and ISO IEC JTC 1/SG 29 joint family standards, all based on conventional hybrid coding framework
  - H.262/MPEG-2 (1995)
  - H.264/MPEG-4 AVC (2003)
  - H.265/MPEG-H HEVC (2013)
  - H.266/MPEG-I VVC (2020)
- JVET NNVC (Neural-Network based Video Coding)
  - VVC is finalized in July 2020
  - NNVC to study and develop coding technologies beyond VVC's capacity using NN based coding tools



JVET NNVC History and Timeline

- **January 2018**: JVET established AHG9 to study NN based video coding tools.
  - More than 100 experts participated the AHG activities and proposed 40 contributions through the study.
  - NN based coding tools in a broad range of technology options were proposed, and coding performance improvement was demonstrated during the two-year study.
- **October 2018**: JVET established evaluation methodology for NNVC (JVET-L1006, JVET-M1006)
- **January 2019**: JVET established Core Experiment on NN based video coding (JVET-M1033, JVET-N1030)
- **June 2020**: JVET re-established AHG11 to study NN based video coding with the goal of developing a potential VVC extension supporting learning-based video coding tools.
- **October 2020**:
  - JVET established a common test condition (CTC) for NNVC (JVET-T2006, JVET-U2016).
  - JVET established Exploration Experiments on NN based video coding (JVET-T2023, JVET-U2023).

JVET NNVC Test Conditions

- Test sequences and conditions (JVET-U2016)
  - Class A~F for VVC development mandatory; class H (HDR) optional
  - VTM-11.0 as anchor
  - All Intra (AI), Random Access (RA) and Low Delay B (LDB) tested
  - CTC for VVC development
  - Config. 1 (hybrid structure): QP@{22, 27, 32, 37, 42}
  - Config. 2 (E2E structure): Each rate points to be within ±10% of the rates of the anchor QP @ {27, 32, 37, 42 }.

Class	Sequence name	Frame count	Frame rate	Bit depth	Intra	Random access	Low-delay
A1	Tango2	294*	60	10	M	M	O*
A1	FoodMarket4	300*	60	10	M	M	O*
A1	Campfire	300*	30	10	M	M	O*
A2	CatRobot	300*	60	10	M	M	O*
A2	DaylightRoad2	300*	60	10	M	M	O*
A2	ParkRunning3	300*	50	10	M	M	O*
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
H2	DayStreet2	300	60	10	O	O	-
H2	FlyingBirds3	300	60	10	O	O	-
H2	PeopleInShoppingCenter2	300	60	10	O	O	-
H2	SunsetBeach3	300	60	10	O	O	-

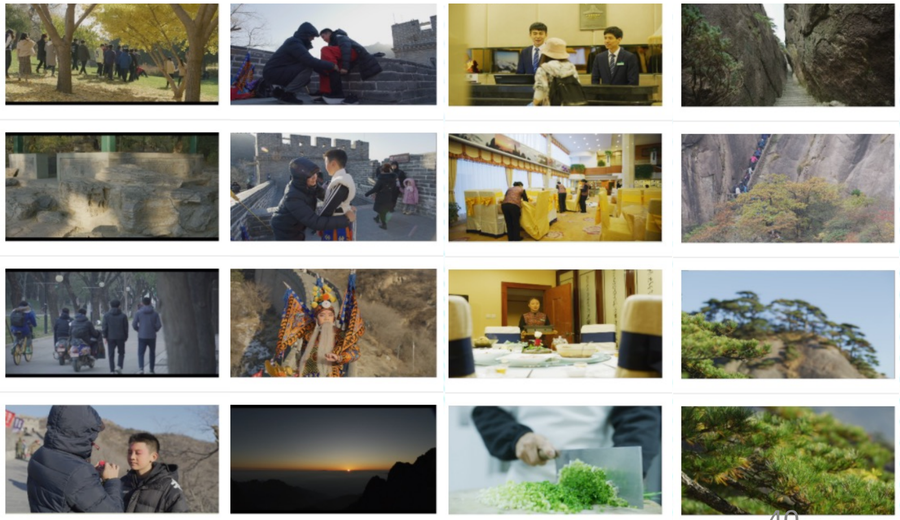
M: Mandatory; O: Optional

S. Liu, A. Segall, E. Alshina, R. Liao, "JVET common test conditions and evaluation procedures for neural network-based video coding technology", JVET-U2016, Jan. 2021.

JVET NNVC Dataset

- Training dataset
  - Candidate videos during VVC development
    - Various resolutions and durations
  - DIV2K (used in CVPR challenge)
    - 1000 images in 2K resolution, RGB format
  - BVI-DVC from University of Bristol
    - 200 different sequences.
    - Four resolutions of each content up to 4K; 65 frames each.
  - Tencent Video Dataset (TVD)
    - 86 different sequences; in 3840x2160 resolution; 65 frames each.
    - X. Xu, et al. "A video dataset for training in neural network based video coding", JVET-U0116, Jan. 2021.
- Training conditions
  - A proposal may elect to use only a subset of the defined sequences if desired.
  - A proposal may also elect to sub-divide the set of defined sequences into different sub-sets.
  - The use of any additional sequences should be described in the contribution

Thumbnails of BVI-DVC videos



Thumbnails of TVD videos

JVET NNVC Complexity Evaluation

- Complexity Measurement
  - Platform information
  - Training information
  - Inference information
  - Encoder/Decoder runtime (CPU, GPU, combined)

Network Information in Inference Stage		
Mandatory	HW environment:	
	GPU Type	CPU only
	Framework:	Pytorch 1.2.0
	Number of GPUs per Task	0
	Total Parameter Number	22371
	Parameter Precision (Bits) in Float	32
	Memory Parameter (MB)	0.085338593
	MAC (Giga) per 3840x2160 pixels	186.64
Optional	Total Conv. Layers	13
	Total FC Layers	0
	Total Memory (MB)	
	Batch size:	1
	Patch size	
	Changes to network configuration or weights required to generate rate points	(e.g.)
	Peak Memory Usage	
	Other information:	

Network Information in Training Stage		
Mandatory	GPU Type	Intel Core i7-8700 3.20 GHz (6 cores), 32GB RAM and one 11 GB Nvidia GTX 1080Ti
	Framework:	PyTorch v1.4.0
	Number of GPUs per Task	1 Nvidia GTX 1080Ti was used for training
	Epoch:	300
	Batch size:	16
	Training time:	48h
	Training data information:	DIV2K still images
	Training configurations for generating compressed training data (if different to VTM CTC):	QP = 22, 27, 32, 37
Optional	Number of iterations	
	Patch size	64x64
	Learning rate:	1.00E-03
	Optimizer:	ADAM
	Loss function:	L1
	Preprocessing:	randomly cropped and rotated
	Other information:	

Y. Li, S. Liu, K. Kawamura, “Methodology and reporting template for neural network coding tool testing”, JVET-M1006, Jan. 2019.

S. Liu, A. Segall, E. Alshina, R. Liao, “JVET common test conditions and evaluation procedures for neural network-based video coding technology”, JVET-U2016, Jan. 2021.

JVET NNVC Tools and Experiments

- In JVET NNVC, the latest progress include both hybrid learning based coding tools, and end-to-end coding tools.
  - In-loop filter
  - Intra prediction
  - Inter-prediction
  - Super resolution
  - End-to-end coding

- Based on input contributions, JVET NNVC established the Exploration Experiments (EE) on NN-based video coding, to investigate commonly interested topics (JVET-T2023, JVET-U2023)

IEEE DCSC FVC

- The Future Video Coding Study Group (FVC-SG) under IEEE Data Compression Standard Committee (DCSC) started investigating deep learning-based image and video compression in 2019.
- **April 2020**: A Call for Evidence (CfE) has been issued. Seven teams registered and submitted results.
- **June 2020**: CfE results have been collected and evaluated.
- **November 2020**: FVC has released the reference software NIC-0.1 as an open-source package (GitHub: https://github.com/fvc-sg/NIC)
- **March 2021**: FVC has released the reference software NIC-0.2 with some updated features.

IEEE DCSC FVC Test Conditions

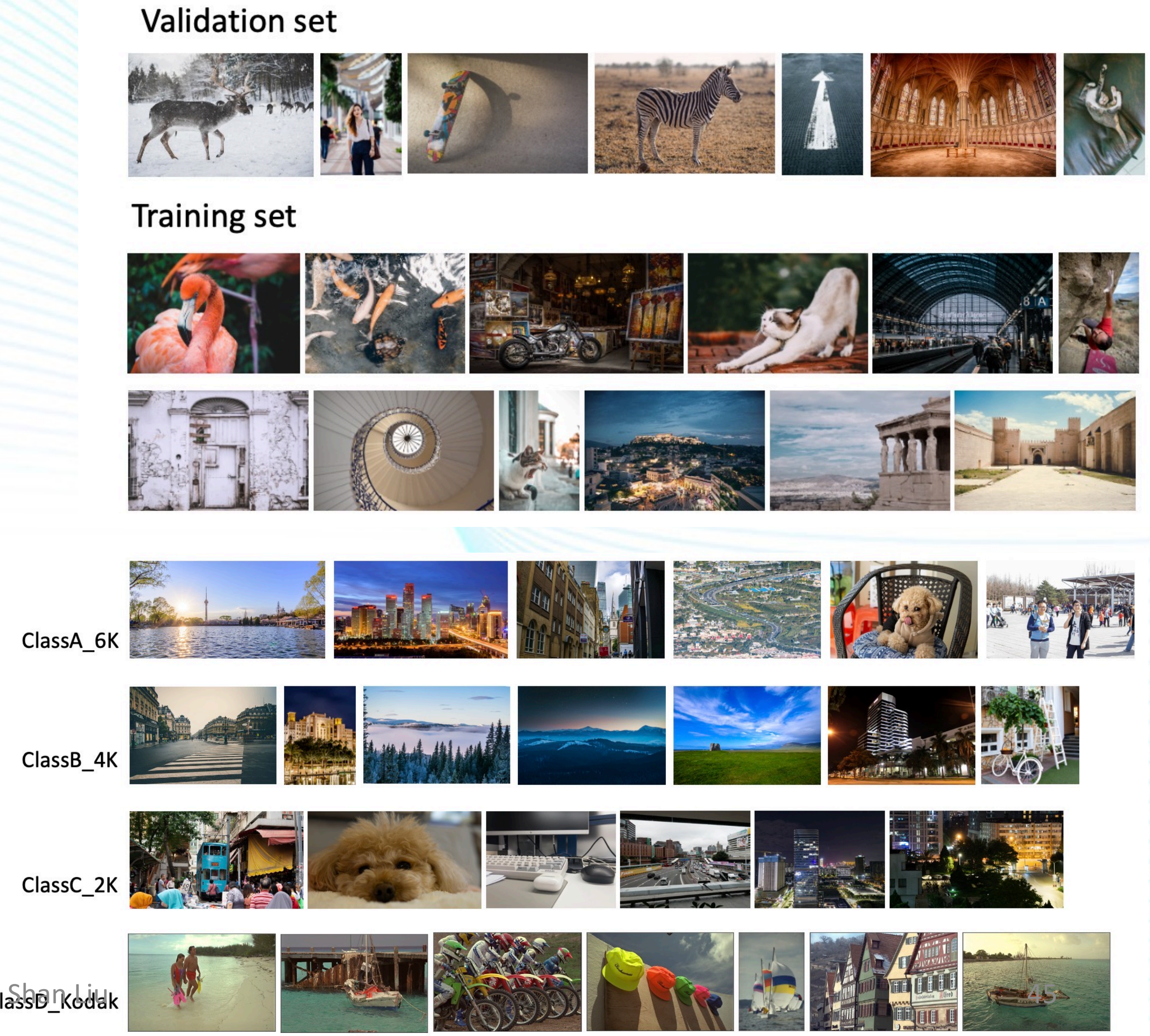
- Anchors
  - BPG (0.9.8), VVC (VTM-8.0 AI config.)
- Target bitrates
  - bpp @{0.06, 0.12, 0.25, 0.5, 1.0, 1.5}
  - Fix-rate model: use λ values in the loss function to train different models for MSE and MS-SSIM criteria
  - Variable-rate model: compress image to meet the 6 target bpp rates with max deviation <=5%
- Evaluation procedure
  - Objective assessment
    - For metrics: YUV 4:2:0 VMAF, RGB PSNR, RGB MS-SSIM (dB), YUV 4:4:4 PSNR, Y MS-SSIM (dB)
  - Subjective assessment
    - MOS from double stimulus protocol
  - Encoder/Decoder runtime
    - processor platforms should be specified

$$L = \lambda * D + R_{main} + R_{hyper}$$

Loss Criteria	λ value			
MSE	0.04	0.08	0.16	0.32
	0.64	1.28	3.20	6.40
MS-SSIM	2	4	8	16
	32	64	128	256

IEEE DCSC FVC Dataset

- NIC_Dataset is an open dataset
  - https://www.bitahub.com/dataset
- Training set
  - 607,714 256x256 patches, cropped from 1,600 original images and the 2x and 4x down-sampled versions
- Validation set
  - 169,798 256x256 patches, cropped from 293 original images and the 2x and 4x down-sampled versions
- Test set
  - 96 images with 4 different resolutions (ClassA_6K, ClassB_4K, ClassC_2K, ClassD_Kodak)



Outline

- Overview of learning-based visual data compression
    - Learning-based image compression
    - Learning-based video compression
    - Learning-based volumetric visual data compression
- Standard activities on learning-based visual data compression
    - JPEG AI
    - JVET NNVC
    - IEEE DCSC FVC
- Discussion and related work

Discussion

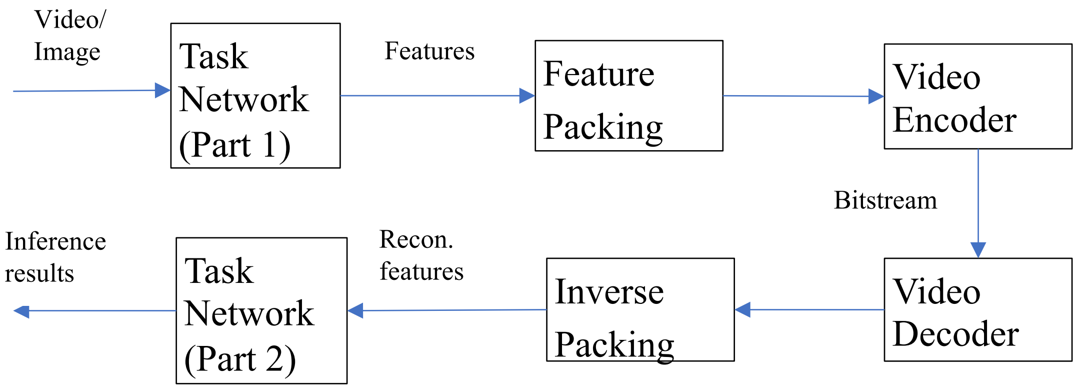
- Complexity issue (JVET-U0023)
  - Decoder runtime typically 30x~400x (with CPU) when compared to conventional solutions
  - Lack of cost-efficient hardware support
- Model limitation
  - Network processing (such as Hyper prior) performs multiple down-sampling to extract features, which impose limitation on feasible block sizes
- Traditional coding method also shows promising results
  - 11% improvement over VVC can be achieved via conventional methods, with ~4x decoder runtime (JVET-U0100)

Related Work (1) Video Coding for Machines

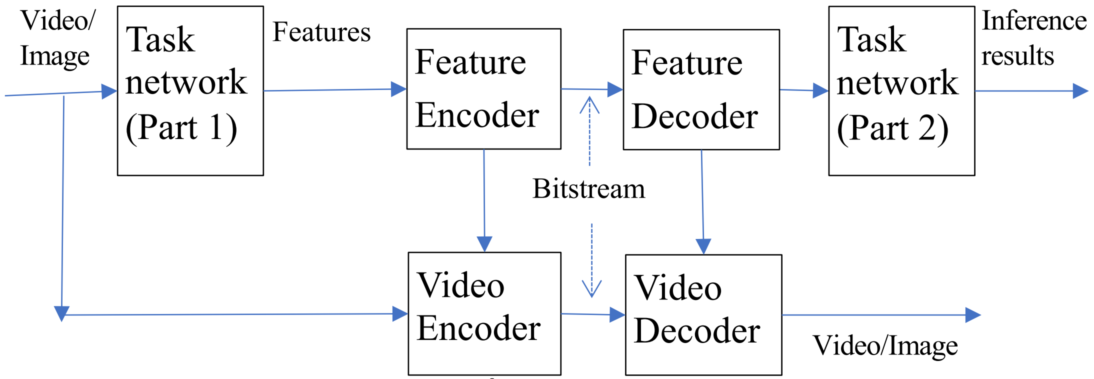
- Due to emerging 5G and IoT technologies, an increased portion of video traffic will be consumed by machines leading to a variety of machine vision tasks such as
  - Object detection/segmentation/tracking, action recognition, etc.
- Machine vision differs from human vision in multiple aspects
  - Such as sensitivity, purpose, and evaluation metric
- Video coding for machines becomes an interesting and challenging problem
  - Traditional video codec may not compress video for machine vision efficiently
- MPEG has created an Ad-Hoc group called “VCM” in July 2019
  - To study use cases, requirements and standardization of VCM technologies
  - Call for Evidence has issued in January 2021

MPEG VCM Processing Pipelines

Pipeline 1



Pipeline 2

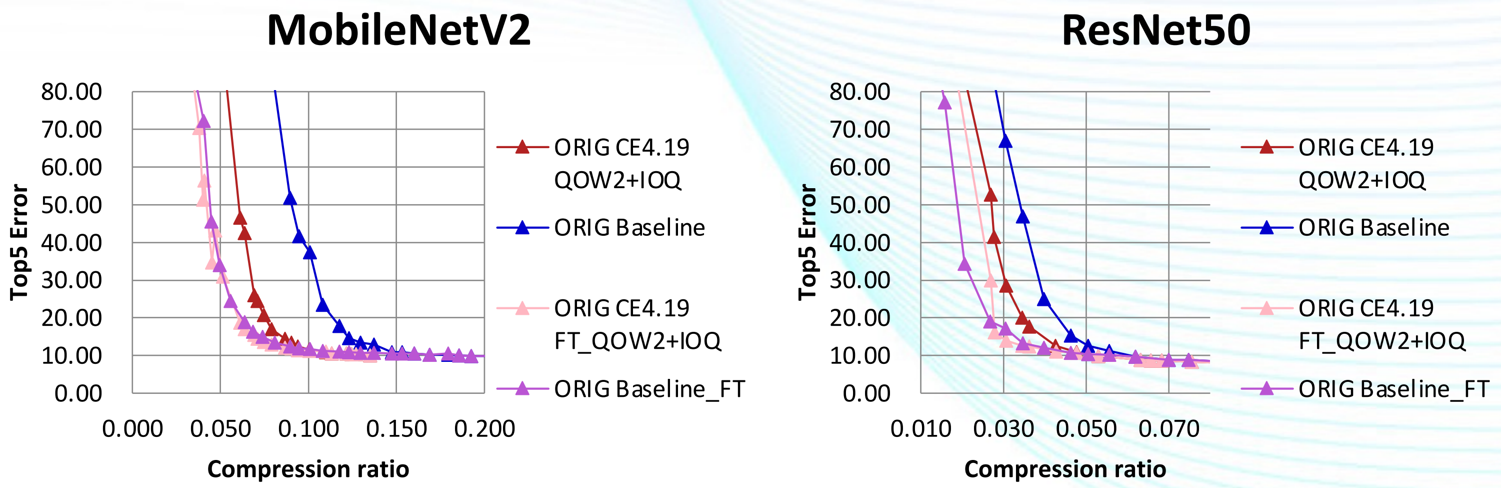


Pipeline 3

M. Rafie, Y. Zhang, S. Liu, "Evaluation Framework for Video Coding for Machines", ISO/IEC JTC 1/SC 29/WG2 output document, N00041, January 2021, Online.

Related Work (2) Neural Network Compression

- MPEG has issued a CfP for Neural network compression in MPEG 125 meeting in January 2019.
  - To reduce model size, and enable model running on low power edge devices
  - The first part of the standardization has been completed (now in DIS).
- MPEG NNR
  - Combined structure pruning and unstructured sparsity
  - Micro-structured sparsity
  - Micro-structured unification
  - Low rank decomposition
  - Local scaling adaption
  - Uniform quantization
  - Dependent quantization
  - Entropy coding



Conclusion

- Overview of learning-based visual data compression
  - Learning-based image compression
  - Learning-based video compression
  - Learning-based volumetric visual data compression

- Standard activities on learning-based visual data compression
  - JPEG AI
  - JVET NNVC
  - IEEE DCSC FVC

- Discussion and related work
  - Video Coding for Machines
  - Neural Network Compression

Q & A

shanl@tencent.com