Supplementary Material: USTC-TD: A Test Dataset and Benchmark for Image and Video Coding in 2020s

Zhuoyuan Li, Junqi Liao, Chuanbo Tang, Haotian Zhang, Yuqi Li, Yifan Bian, Xihua Sheng, Xinmin Feng, Yao Li, Changsheng Gao, Li Li, *Member, IEEE,* Dong Liu, *Senior Member, IEEE,* and Feng Wu, *Fellow, IEEE*

I. Overview of the Evaluative Advanced Image/Video Compression Schemes

Here we mention the characteristics of each test image/video evaluative compression scheme in our experiments (Section V of main text). The test instruction of each method is released on the open-sourced website: https://github.com/EsakaK/USTC-TD in detail.

1) Standardized Schemes: ***BPG*** [1]: *BPG* (Better Portable Graphics) is an image format. Its purpose is to replace the *JPEG* image format when quality or file size is an issue. It supports the same chroma formats as *JPEG* and has the higher compression ratio than *JPEG*.

High Efficiency Video Coding (H.265/HEVC) [1]: High Efficiency Video Coding (*HEVC*), also known as *H.265*. Compared to *H.264/AVC*, *H.265/HEVC* offers from 25% to 50% better data compression at the same level of video quality, or substantially improved video quality at the same bit rate.

Versatile Video Coding (H.266/VVC) [2]: Versatile Video Coding (*H.266/VVC*) is a video compression standard finalized on 2020, standardized by the Joint Video Experts Team (JVET), VCEG working group of ITU-T Study Group 16 and the MPEG working group of ISO/IEC JTC 1/SC 29. It is the successor to *H.265/HEVC*. It improves compression performance and supports for a very broad range of applications.

IEEE 1857.11 [2]: *IEEE 1857.11* provides efficient, neural network-based coding tools for compression, decompression, and facilitate the compression and decompression on top of neural network-oriented computing infrastructures like neural network processing units (NPUs).

2) Learned Image Compression Schemes: ***Factorized Model*** [3]: *Factorized Model* introduces the convolutional-based transform network employing generalized divisive normalization [4] and the factorized entropy model. It suggests the additive uniform noise to address non-differentiable quantization. This is the first learned image compression method that surpasses *JPEG2000* [5] on RGB BD-rate *PSNR/MS-SSIM*.

Hyperprior Model [3]: The pioneering work proposes a hyperprior entropy model for learned image compression, significantly enhancing compression performance. It has since been widely used and remains influential in the field.

Autoregressive Model [6]: *Autoregressive Model* proposes combining an autoregressive spatial context model with a hyperprior for more accurate entropy estimation. This method marks the first learning-based image compression approach to outperform *BPG* on RGB BD-rate *PSNR/MS-SSIM*.

Cheng2020 [7]: *Cheng2020* utilizes the discretized gaussian mixture model to estimate the distributions of latent codes, while employing attention modules in transform to enhance performance. Notably, this is the first work to achieve comparable performance with the latest compression standard *VVC* [2] in terms of RGB BD-rate *PSNR/MS-SSIM*.

iWave++ [8]: *iWave++* proposes a versatile learned image compression scheme with a trained wavelet-like transform. It supports lossy/lossless image compression. This model is accepted as the high-profile model in *IEEE 1857.11*.

ELIC [9]: *ELIC* adopts the space-channel context model in entropy estimation and stacked residual blocks as the non-linear transform. This method not only demonstrates superior performance but also supports fast preview decoding.

MLIC++ [10]: *MLIC++* utilizes channel-wise, local spatial, and global spatial information to achieve better entropy estimation. Compared to concurrent models like *ELIC*, this method achieves state-of-the-art compression performance.

3) Learned Video Compression Schemes: ***DVC_Pro*** [11], [12]: The first work proposes the motion-compensated prediction and residual coding framework of end-to-end learned video compression. Under intra period 10, experimental results show that *DVC_pro* performs better than *H.264/AVC* [13] on RGB BD-rate *PSNR/MS-SSIM*. Besides, *DVC_pro* achieves comparable performance with *H.265/HEVC* [1] on RGB BD-rate *MS-SSIM*.

DCVC [14]: The first work proposes the motion-compensated prediction and conditional coding framework of learned video compression. *DCVC* utilizes the learned temporal correlation between the current frame and the predicted frame rather than the subtraction-based residual. Under intra period 10, it performs better than *DVC_Pro* [12] and x265 on RGB BD-rate *PSNR/MS-SSIM*. Specifically, it achieves an average 18.40% and 17.82% RGB BD-rate reduction, in terms of *PSNR* compared to x265 and *DVC_Pro* [12].

CANF-VC [15]: *CANF-VC* is the first conditional augmented normalizing flows-based end-to-end learned video compression framework. *CANF-VC* leverages the conditional augmented normalizing flows to learn a video generative

Date of current version March 16, 2025. *(Z. Li and J. Liao contributed equally to this work.) (Corresponding authors: Dong Liu and Feng Wu.)*

The authors are with the MOE Key Laboratory of Brain-Inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230027, China (e-mail: {zhuoyuanli, liaojq, cbtang, zhanghaotian, lyq010303, esakak, xhsheng, xmfeng2000, mrliyao}@mail.ustc.edu.cn; {changshenggao, lil1, dongeliu, fengwu}@ustc.edu.cn).

[1]Available online at https://bellard.org/bpg.

[2]Available online at https://sagroups.ieee.org/fvc/.

model for conditional inter-frame coding, and extends the conditional coding to motion coding, forming a purely conditional coding framework. Under intra period 10/12, its experimental results perform better than [16], [14] and [1] on RGB BD-rate *PSNR/MS-SSIM*. Under intra period 32, its experimental results perform better than those of [16], [14] on RGB BD-rate *PSNR*, and H.265 [1] on RGB BD-rate *MS-SSIM*.

TCM-VC [17]: Based on *DCVC* [14], *TCM-VC* further proposes the multi-scale temporal context mining to better utilize the temporal correlation. Under intra period 32, the experimental results show that the compression performance of neural video codecs [11], [12], [16], [18], [14] is greatly reduced while their framework still achieves an average 14.4% RGB BD-rate reduction against HM-IPP on *PSNR* and achieves about 21.1% BD-rate reduction on *MS-SSIM*.

DCVC-HEM [19]: Based on *TCM-VC* [17], *HEM* further designs a parallel-friendly entropy model that explores both temporal and spatial dependencies. Besides, it also supports variable bitrates in a single mode. *HEM* is the first end-to-end neural video codec to exceed *H.266/VVC* [20] using the highest compression ratio configuration. Under intra period 32, the experimental results show that the compression performance of neural video codecs [12], [16], [18], [14], [17] is greatly reduced while their framework still achieves an average 4.7% RGB BD-rate reduction against VTM-IPP in terms of *PSNR* and achieves about 46.4% BD-rate reduction in terms of *MS-SSIM*.

OOFE [21]: *OOFE* proposes an offline and online optical enhancement strategy for the flow-based end-to-end learned video compression framework, which is integrated into *DCVC* and *DCVC-DC*, respectively. Experimental results demonstrate that the proposed offline and online enhancement together achieves on average 13.4% bitrate saving for *DCVC* [14] and 4.1% bitrate saving for *DCVC-DC* [22] on RGB BD-rate *PSNR* when intra period is 12.

VNVC [23]: The first work proposes a versatile neural video coding framework for both human and machine vision. It reports the experimental results on video reconstruction, enhancement, analysis tasks. For video reconstruction, it regards *H.266/VVC* official reference software VTM-13.2 with VTM-IPP configuration (one reference frame and flat QP) as the anchor. Under intra period 12, it achieves an average 7.0% BD-rate reduction against VTM-IPP in terms of *PSNR* and achieves about an average 49.9% BD-rate reduction in terms of *MS-SSIM*. Under intra period 32, the results show that the performance of neural video codecs [12], [14], [15], [17] is greatly reduced while their framework still achieves an average 4.6% BD-rate reduction against VTM-IPP on *PSNR* and achieves about 50.3% BD-rate reduction on *MS-SSIM*.

SDD [24]: Based on the motion-compensated prediction and conditional coding framework, *SDD* proposes a structure and detail decomposition-based motion model and a long short-term temporal contexts fusion mechanism. Under intra period 32, the experimental results show that the compression performance of neural video codecs [11], [12], [16], [18], [14], [15], [17], [19] is greatly reduced while their framework still achieves an average 13.4% RGB BD-rate reduction against VTM-IPP in terms of *PSNR* and achieves about 48.0% BD-rate reduction in terms of *MS-SSIM*.

DCVC-DC [22]: Based on *HEM* [19], *DCVC-DC* further increases the context diversity in both temporal and spatial dimensions by introducing the group-based offset diversity and quadtree-based partition. Under intra period 32, the experimental results show that the compression performance of neural video codecs [11], [12], [14], [17], [19] is greatly reduced while their framework still achieves an average 17.8% RGB BD-rate reduction against VTM-IPP on *PSNR* and achieves about 47.6% BD-rate reduction on *MS-SSIM*.

DCVC-FM [25]: Based on *DCVC-DC* [22], *DCVC-FM* exploits the training with longer video and proposes a periodically refreshing mechanism. Besides, it modulates the latent feature via the learnable quantization scaler to support a wide quality range in a single model. Under intra period 32, the experimental results show that the compression performance of neural video codecs [11], [12], [18], [14], [15], [17], [19], [22] while their framework still achieves an average 20.3% RGB BD-rate reduction against VTM-IPP on *PSNR*. Under intra period 96, *DCVC-FM* can also achieve on average 25.3% and 38.3% YUV BD-rate reduction, respectively, against VTM-IPP and DCVC-DC on *PSNR*.

II. The Benchmark of Implicit Neural Representation (INR)-based Video Compression Schemes on USTC-TD

In this section, first, we present the experimental configurations employed for the evaluation of implicit neural representation (INR)-based video compression schemes. Second, we evaluate the recent advanced INR-based video compression schemes on USTC-TD under different metrics, and benchmark their performance on USTC-TD. Third, we analyze the benchmarked performance.

A. Experimental Settings

1) Selection of Evaluative Compression Schemes: For advanced INR-based schemes, we select the *NeRV* [26], *HNeRV* [27], and *HiNeRV* [28]. Since the *NVRC* [29] has not been open-sourced yet, we will update the results on open-sourced website once it is released.

2) Overview of the Evaluative INR-based Video Compression Schemes: For ***NeRV*** [26], *NeRV* is a novel neural representation for videos that encodes videos into neural networks, treating them as implicit functions rather than frame sequences. Unlike conventional video compression methods that rely on complex pipelines, *NeRV* directly maps frame indices to RGB frames using a neural network. This approach significantly improves efficiency, achieving 25× to 70× faster encoding speeds and 38× to 132× faster decoding speeds compared to pixel-wise implicit representations. *NeRV* also enables video compression by reformulating it as a model compression problem, leveraging pruning, quantization, and entropy encoding to achieve competitive performance with traditional codecs like *H.264/AVC* and *H.265/HEVC*. For ***HNeRV*** [27], *HNeRV* is a hybrid neural representation for videos that improves upon implicit methods like *NeRV* by introducing content-adaptive embeddings and a redesigned decoder architecture.

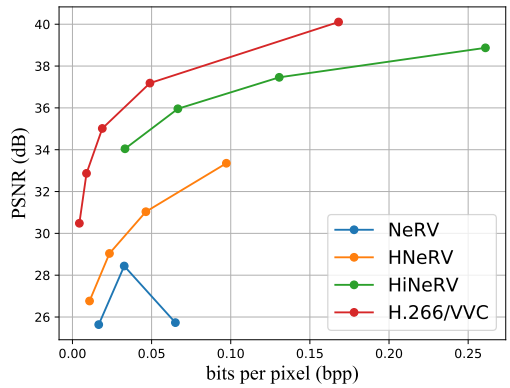


Fig. 1. Overall rate-distortion (RD) curves of advanced INR-based compression schemes on *PSNR* metric.

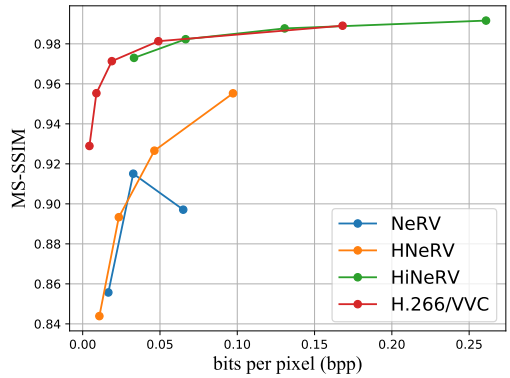


Fig. 2. Overall rate-distortion (RD) curves of advanced INR-based compression schemes on *MS-SSIM* metric.

Unlike *NeRV*, which relies on fixed positional embeddings, *HNeRV* employs a learnable encoder to generate content-aware embeddings, enhancing video regression capacity and enabling better frame interpolation. In addition, the *HNeRV* block architecture redistributes model parameters more evenly, improving reconstruction quality (+4.7 *PSNR*) and convergence speed (16× faster). *HNeRV* retains the advantages of neural representations for video compression while also excelling in video inpainting and random-access decoding, making it a flexible and efficient video representation. For ***HiNeRV*** [28], *HiNeRV* is a hierarchical encoding-based neural representation designed for video compression, significantly improving upon previous INR methods. It introduces lightweight layers and hierarchical positional encodings, leveraging depth-wise convolutions, MLPs, and interpolation layers to enhance representation capacity while maintaining efficiency. Unlike conventional INRs that rely on Fourier-based encodings, *HiNeRV* employs multi-resolution local feature grids, enabling faster convergence and higher reconstruction quality. *HiNeRV* also unifies frame-wise and patch-wise video representations, allowing flexible encoding with minimal boundary effects. In addition, it refines the model compression pipeline, incorporating adaptive pruning and quantization-aware training to improve rate-distortion performance. Experimental results on the *UVG* and *MCL-JCV* datasets show that *HiNeRV* outperforms existing INR-based codecs (e.g., 72.3% BD-rate saving over *HNeRV*) and achieves competitive performance compared to learning-based codecs such as *DCVC* and x265 (veryslow). Notably, *HiNeRV* is the first INR-based codec to surpass x265 (veryslow) in the performance of *PSNR* measure, demonstrating the potential of INRs in video compression applications.

3) Testing Configurations and Evaluative Metric of INR-based Video Compression Schemes: For all models tested, we set the number of training epochs to 1200 and batch size (in video frames) to 1. We use the same optimizer as in [26], and employ the same learning objectives for all *NeRV*-based schemes as in the original literature [26]–[28], [30]–[32]. For the evaluative metric, we follow the same setting of *PSNR* and *MS-SSIM* as mentioned in Section V.A (7) of main text, with both metrics calculated and compared in the RGB color space.

B. Experimental Results

In this subsection, we establish the baselines and benchmark the performance of advanced INR-based video compression schemes on the USTC-TD video dataset, and further analyze their performance. Taking bpp as the horizontal axis and the reconstructed *PSNR*/*MS-SSIM* as the vertical axis, we present the rate and distortion curves of different INR-based video compression schemes over the USTC-TD long video dataset in Fig. 1 and 2. From the overall results of *PSNR* metric, we can find that the performance of the advanced traditional schemes far surpasses that of all advanced INR-based schemes on *PSNR* metric. In contrast, for the *MS-SSIM* metric, *HiNeRV* [28] stands out with competitive performance against advanced traditional schemes and significantly surpasses other INR-based approaches.

III. Specific Performance of Objective Quality Evaluation of Evaluative Advanced Image/Video Compression Schemes on USTC-TD

Here we supply the specific rate-distortion (RD) curves of advanced image/video compression schemes on USTC-TD dataset under different metrics (Section V.B of main text). The results of USTC-TD image dataset are mentioned in Fig. 3, 4 for *PSNR* metric, and 5, 6 for *MS-SSIM* metric, and 7, 8 for *VMAF (MSE model)* metric, and 9, 10 for *VMAF (MS-SSIM)* metric. The results of USTC-TD video dataset are mentioned in Fig. 11, 12, 13, 14 for short setting with intra period = 32, Fig. 15, 16, 17, 18 for long setting with intra period = 32, Fig. 19, 20, 21, 22 for long setting with intra period = -1,.

IV. Specific Performance of Subjective Quality Evaluation of Evaluative Advanced Image/Video Compression Schemes on USTC-TD

Here we supply the specific MOS results of compressed images/videos of different image/video compression schemes over USTC-TD image/video datasets (Section V.B of main text). The results of USTC-TD image dataset is mentioned in Fig. 23. The results of USTC-TD video dataset is mentioned in Fig. 24.

V. Copyright

The released images and sequences are captured and processed by the University of Science and Technology of China

(USTC). All intellectual property rights remain with USTC. The following uses are allowed for the contributed dataset: (1) Data (images and videos) may be published in research papers, technical reports, and development events. (2) Data (images and videos) may be utilized for standardization activities (e.g., ITU, MPEG, AVS, VQEG). The following uses are NOT allowed for the contributed dataset: (1) Do not publish snapshots in product brochures. (2) Do not use video for marketing purposes. (3) Redistribution is not permitted. (4) Do not use it in television shows, commercials, or movies.

REFERENCES

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021.

[3] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," *arXiv preprint arXiv:1802.01436*, 2018.

[4] J. Ballé, V. Laparra, and E. P. Simoncelli, "Density modeling of images using a generalized normalization transformation," *arXiv preprint arXiv:1511.06281*, 2015.

[5] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal processing magazine*, vol. 18, no. 5, pp. 36–58, 2001.

[6] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," *Advances in Nneural Information Processing Systems*, vol. 31, 2018.

[7] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learned image compression with discretized gaussian mixture likelihoods and attention modules," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7939–7948.

[8] H. Ma, D. Liu, N. Yan, H. Li, and F. Wu, "End-to-end optimized versatile image compression with wavelet-like transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 3, pp. 1247–1263, 2020.

[9] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, "Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5718–5727.

[10] W. Jiang, J. Yang, Y. Zhai, P. Ning, F. Gao, and R. Wang, "Mlic: Multi-reference entropy model for learned image compression," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 7618–7627.

[11] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao, "Dvc: An end-to-end deep video compression framework," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 006–11 015.

[12] G. Lu, X. Zhang, W. Ouyang, L. Chen, Z. Gao, and D. Xu, "An end-to-end learning framework for video compression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3292–3308, 2020.

[13] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[14] J. Li, B. Li, and Y. Lu, "Deep contextual video compression," *Advances in Neural Information Processing Systems*, vol. 34, pp. 18 114–18 125, 2021.

[15] Y.-H. Ho, C.-P. Chang, P.-Y. Chen, A. Gnutti, and W.-H. Peng, "Canf-vc: Conditional augmented normalizing flows for video compression," in *European Conference on Computer Vision*. Springer, 2022, pp. 207–223.

[16] J. Lin, D. Liu, H. Li, and F. Wu, "M-lvc: Multiple frames prediction for learned video compression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3546–3554.

[17] X. Sheng, J. Li, B. Li, L. Li, D. Liu, and Y. Lu, "Temporal context mining for learned video compression," *IEEE Transactions on Multimedia*, vol. 25, pp. 7311–7322, 2022.

[18] R. Yang, F. Mentzer, L. Van Gool, and R. Timofte, "Learning for video compression with recurrent auto-encoder and recurrent probability model," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 2, pp. 388–401, 2020.

[19] J. Li, B. Li, and Y. Lu, "Hybrid spatial-temporal entropy modelling for neural video compression," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 1503–1511.

[20] B. Bross, J. Chen, J.-R. Ohm, G. J. Sullivan, and Y.-K. Wang, "Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC)," *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463–1493, 2021.

[21] C. Tang, X. Sheng, Z. Li, H. Zhang, L. Li, and D. Liu, "Offline and online optical flow enhancement for deep video compression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 6, 2024, pp. 5118–5126.

[22] J. Li, B. Li, and Y. Lu, "Neural video compression with diverse contexts," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 616–22 626.

[23] X. Sheng, L. Li, D. Liu, and H. Li, "Vnvc: A versatile neural video coding framework for efficient human-machine vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[24] ——, "Spatial decomposition and temporal fusion based inter prediction for learned video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[25] J. Li, B. Li, and Y. Lu, "Neural video compression with feature modulation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 26 099–26 108.

[26] H. Chen, B. He, H. Wang, Y. Ren, S. N. Lim, and A. Shrivastava, "Nerv: Neural representations for videos," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21 557–21 568, 2021.

[27] H. Chen, M. Gwilliam, S.-N. Lim, and A. Shrivastava, "Hnerv: A hybrid neural representation for videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 10 270–10 279.

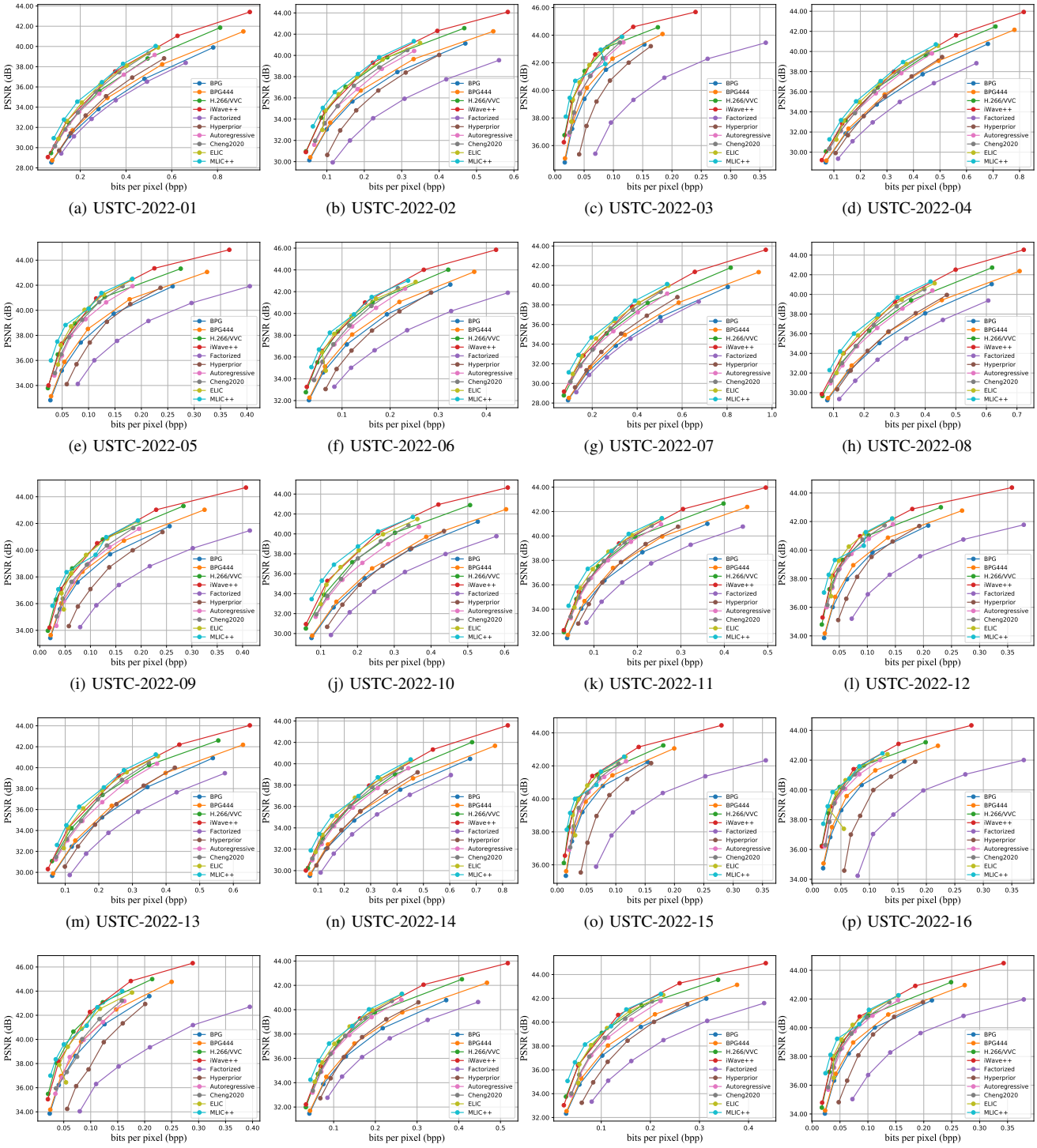
[28] H. M. Kwan, G. Gao, F. Zhang, A. Gower, and D. Bull, "Hinerv: Video compression with hierarchical encoding-based neural representation," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[29] ——, "NVRC: Neural video representation compression," *arXiv preprint arXiv:2409.07414*, 2024.

[30] Z. Li, M. Wang, H. Pi, K. Xu, J. Mei, and Y. Liu, "E-nerv: Expedite neural video representation with disentangled spatial-temporal context," in *European Conference on Computer Vision*. Springer, 2022, pp. 267–284.

[31] Y. Bai, C. Dong, C. Wang, and C. Yuan, "Ps-nerv: Patch-wise stylized neural representations for videos," in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 41–45.

[32] J. C. Lee, D. Rho, J. H. Ko, and E. Park, "Ffnerv: Flow-guided frame-wise neural representations for videos," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 7859–7870.



(a) USTC-2022-01 (b) USTC-2022-02 (c) USTC-2022-03 (d) USTC-2022-04

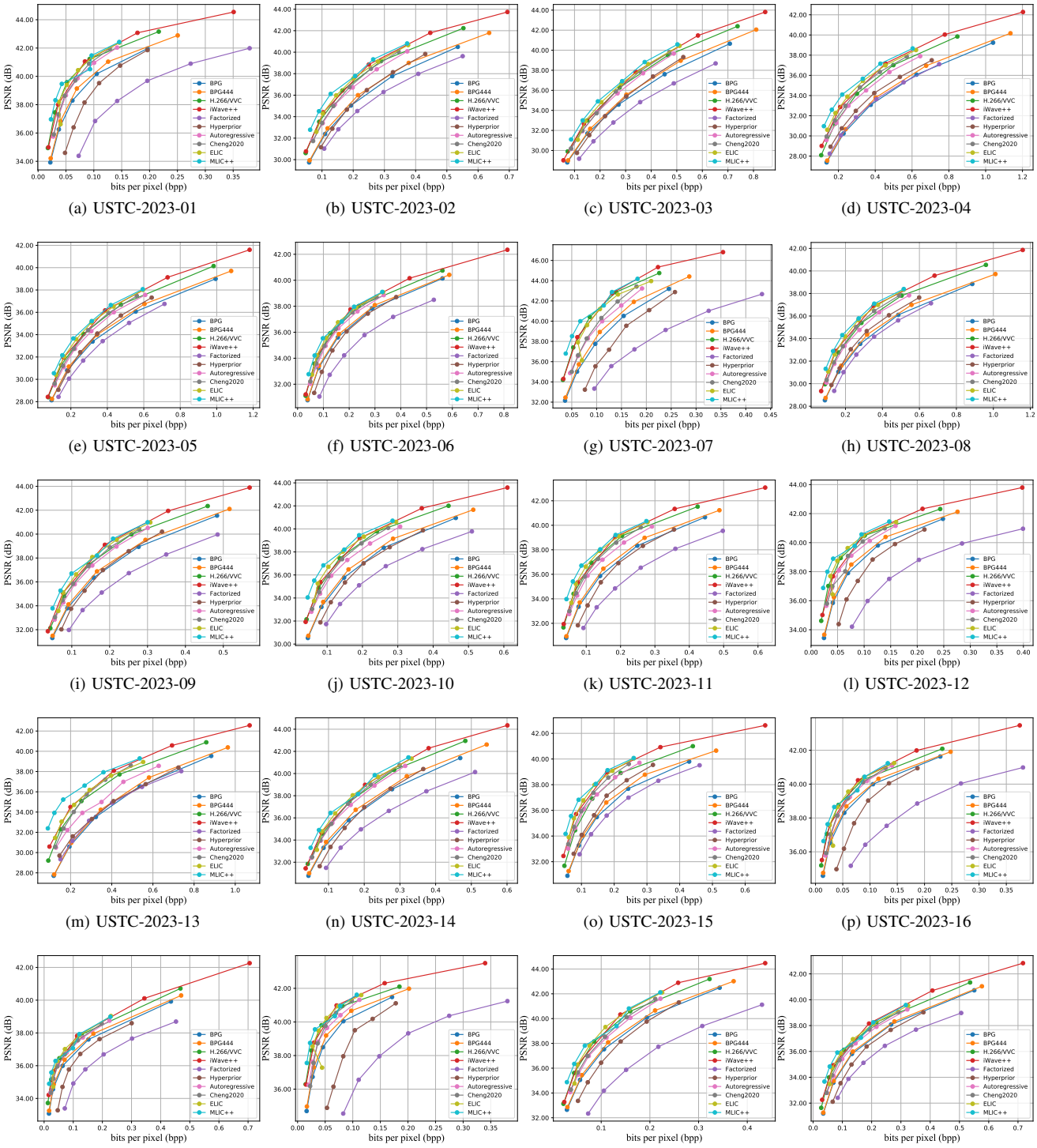
(e) USTC-2022-05 (f) USTC-2022-06 (g) USTC-2022-07 (h) USTC-2022-08

(i) USTC-2022-09 (j) USTC-2022-10 (k) USTC-2022-11 (l) USTC-2022-12

(m) USTC-2022-13 (n) USTC-2022-14 (o) USTC-2022-15 (p) USTC-2022-16

(q) USTC-2022-17 (r) USTC-2022-18 (s) USTC-2022-19 (t) USTC-2022-20

Fig. 3. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2022 under *PSNR* metric.



(a) USTC-2023-01 (b) USTC-2023-02 (c) USTC-2023-03 (d) USTC-2023-04

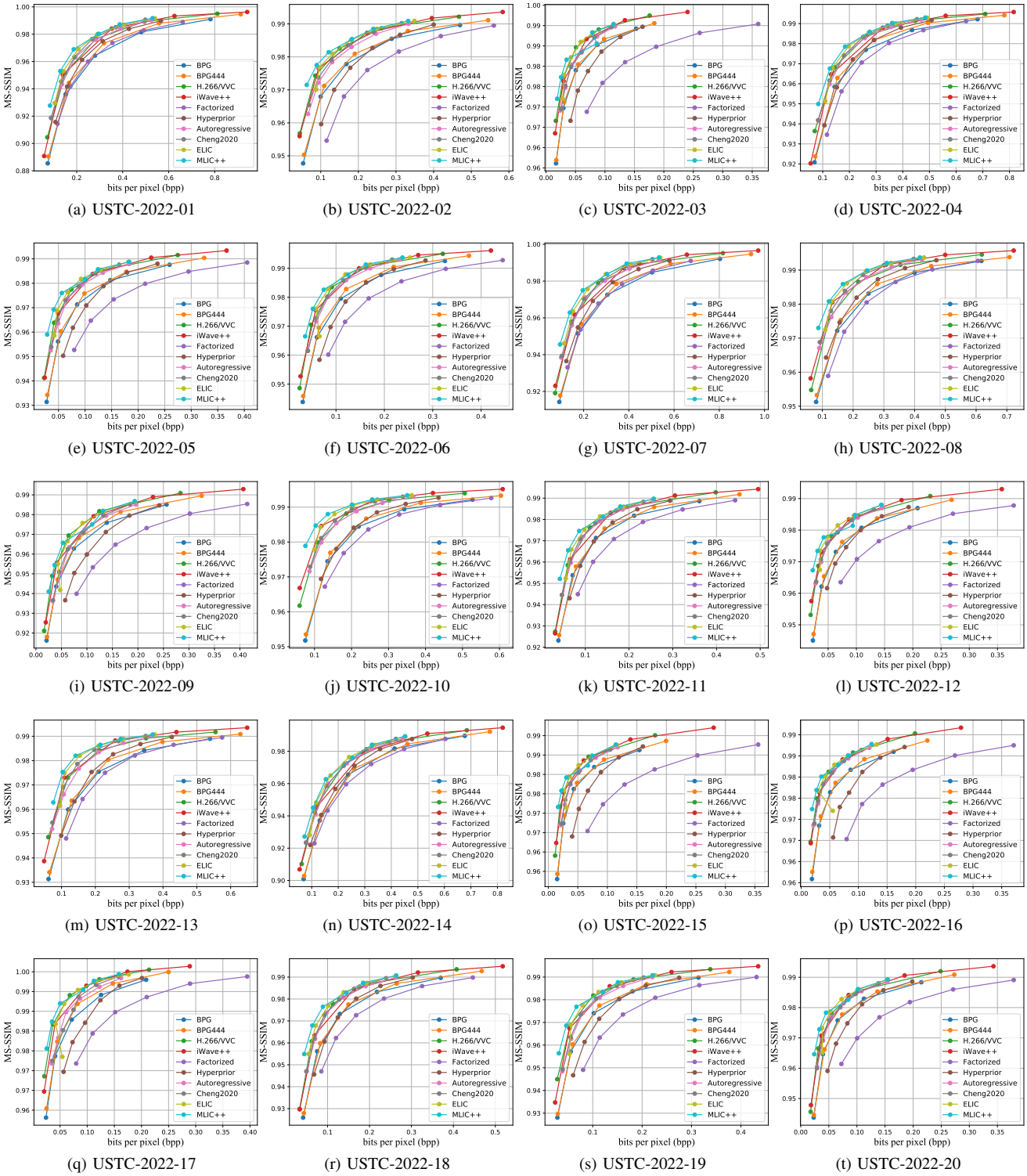
(e) USTC-2023-05 (f) USTC-2023-06 (g) USTC-2023-07 (h) USTC-2023-08

(i) USTC-2023-09 (j) USTC-2023-10 (k) USTC-2023-11 (l) USTC-2023-12

(m) USTC-2023-13 (n) USTC-2023-14 (o) USTC-2023-15 (p) USTC-2023-16

(q) USTC-2023-17 (r) USTC-2023-18 (s) USTC-2023-19 (t) USTC-2023-20

Fig. 4. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2023 under *PSNR* metric.



(a) USTC-2022-01 (b) USTC-2022-02 (c) USTC-2022-03 (d) USTC-2022-04

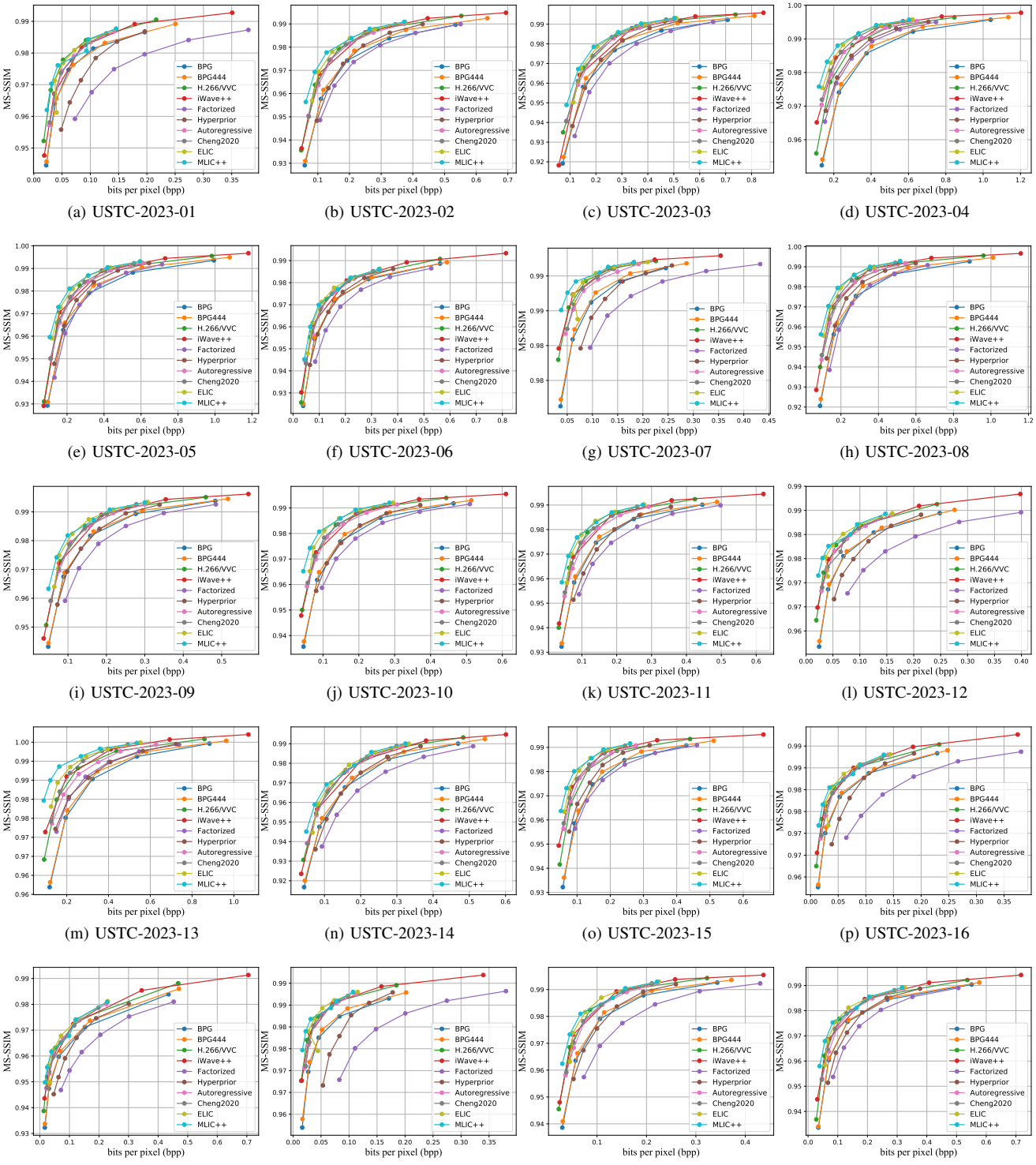
(e) USTC-2022-05 (f) USTC-2022-06 (g) USTC-2022-07 (h) USTC-2022-08

(i) USTC-2022-09 (j) USTC-2022-10 (k) USTC-2022-11 (l) USTC-2022-12

(m) USTC-2022-13 (n) USTC-2022-14 (o) USTC-2022-15 (p) USTC-2022-16

(q) USTC-2022-17 (r) USTC-2022-18 (s) USTC-2022-19 (t) USTC-2022-20

Fig. 5. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2022 under *MS-SSIM* metric.



(a) USTC-2023-01 (b) USTC-2023-02 (c) USTC-2023-03 (d) USTC-2023-04

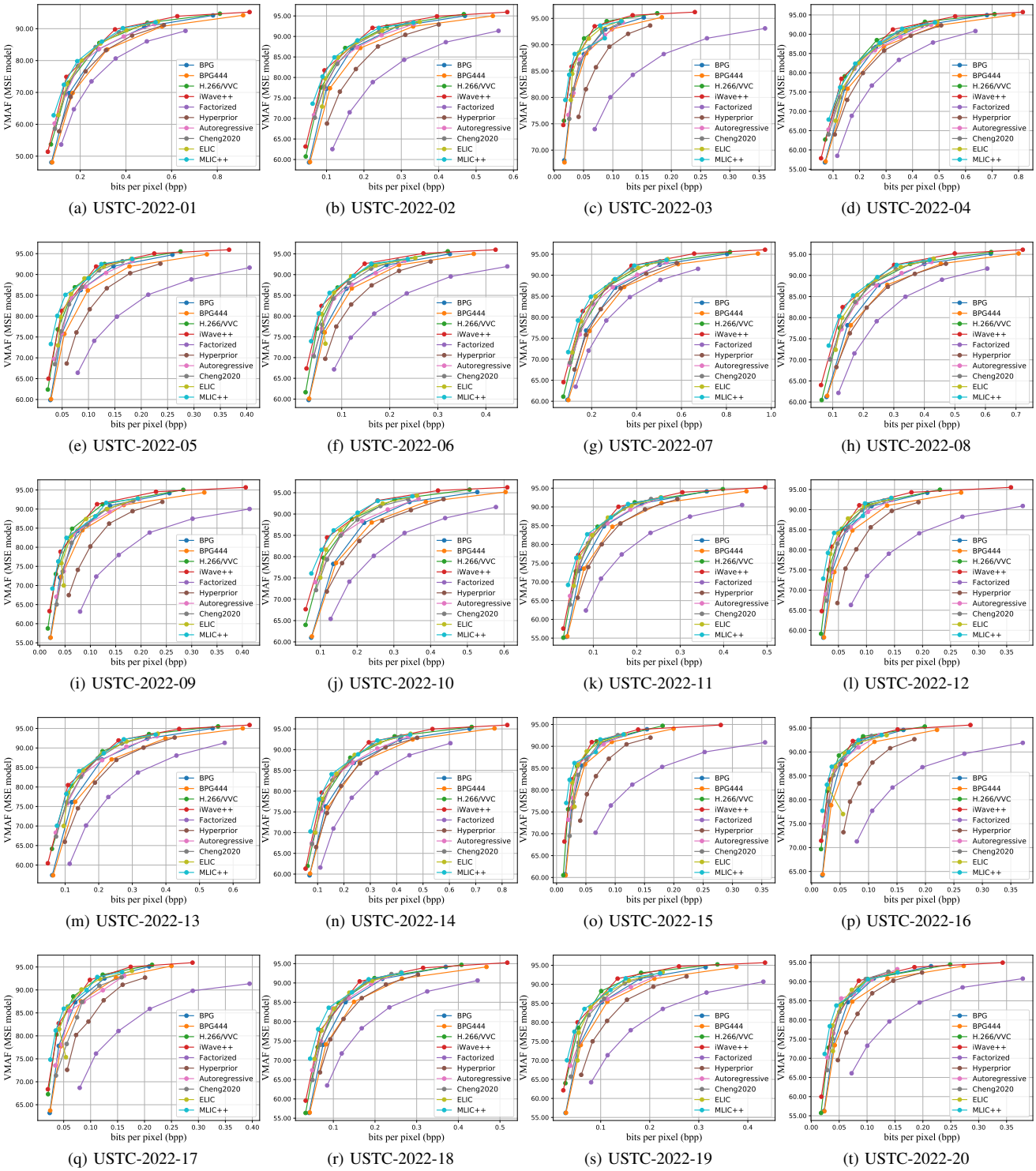
(e) USTC-2023-05 (f) USTC-2023-06 (g) USTC-2023-07 (h) USTC-2023-08

(i) USTC-2023-09 (j) USTC-2023-10 (k) USTC-2023-11 (l) USTC-2023-12

(m) USTC-2023-13 (n) USTC-2023-14 (o) USTC-2023-15 (p) USTC-2023-16

(q) USTC-2023-17 (r) USTC-2023-18 (s) USTC-2023-19 (t) USTC-2023-20

Fig. 6. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2023 under *MS-SSIM* metric.



(a) USTC-2022-01 (b) USTC-2022-02 (c) USTC-2022-03 (d) USTC-2022-04

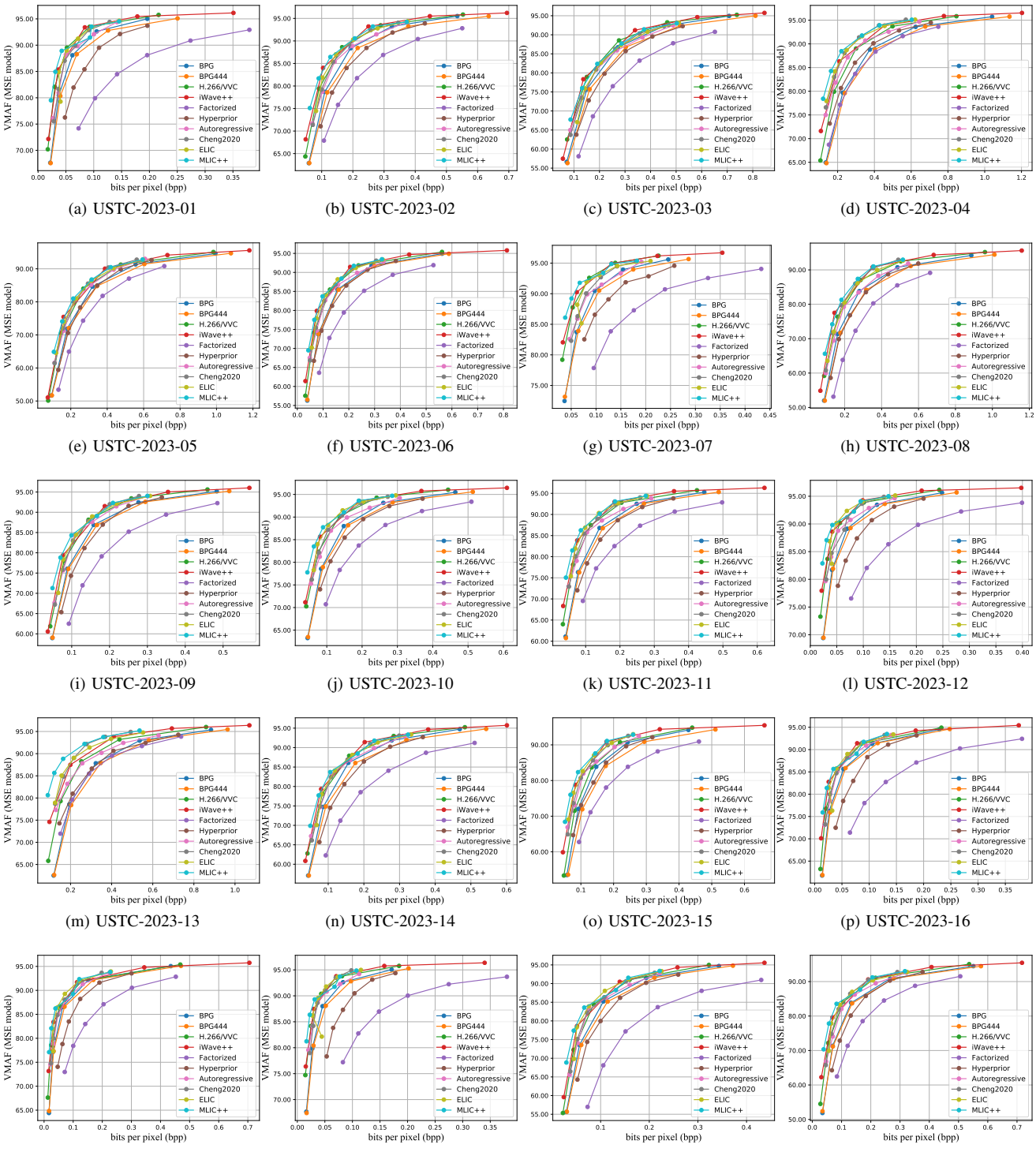
(e) USTC-2022-05 (f) USTC-2022-06 (g) USTC-2022-07 (h) USTC-2022-08

(i) USTC-2022-09 (j) USTC-2022-10 (k) USTC-2022-11 (l) USTC-2022-12

(m) USTC-2022-13 (n) USTC-2022-14 (o) USTC-2022-15 (p) USTC-2022-16

(q) USTC-2022-17 (r) USTC-2022-18 (s) USTC-2022-19 (t) USTC-2022-20

Fig. 7. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2022 under *VMAF (MSE model)* metric.



(a) USTC-2023-01 (b) USTC-2023-02 (c) USTC-2023-03 (d) USTC-2023-04

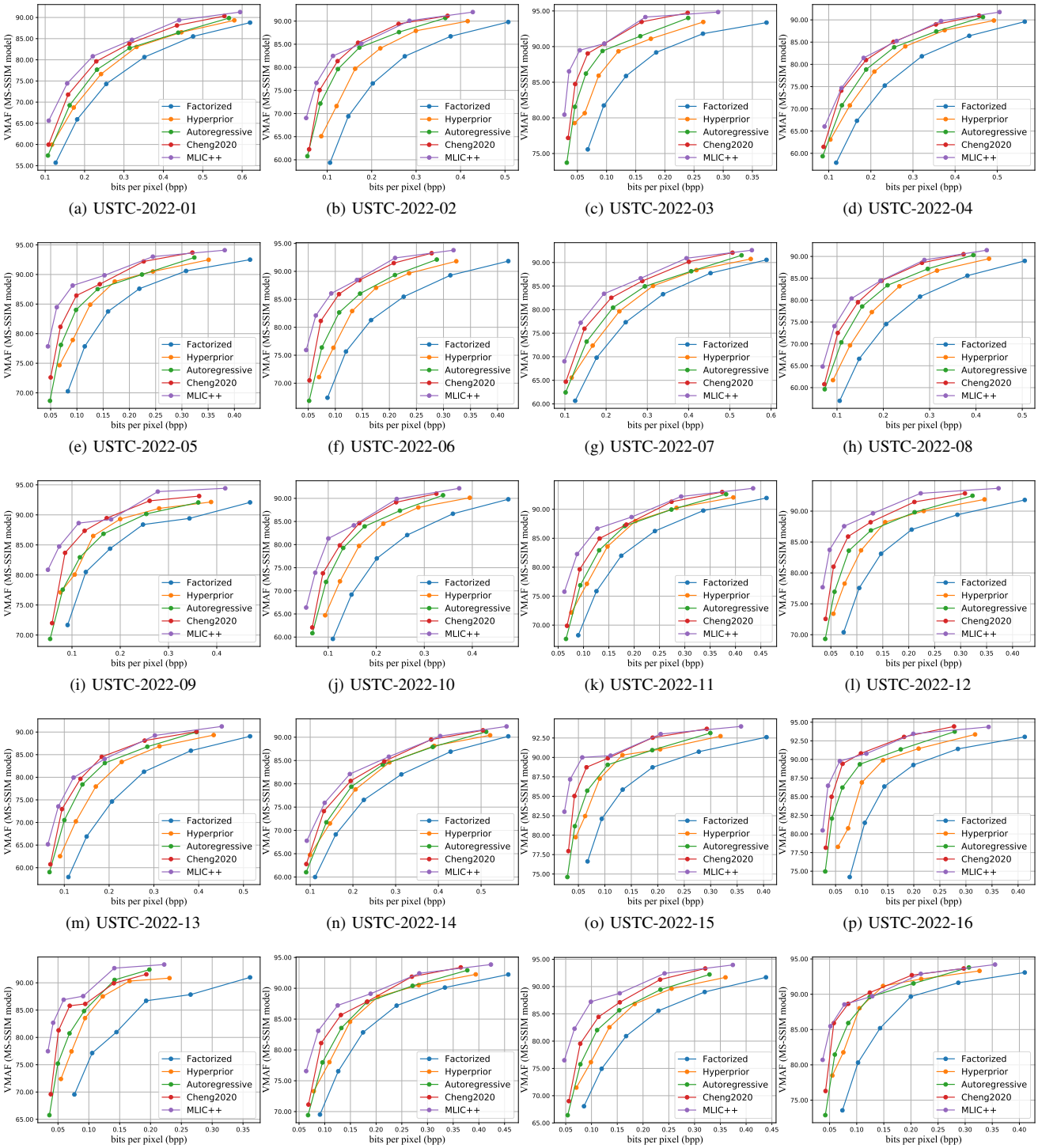
(e) USTC-2023-05 (f) USTC-2023-06 (g) USTC-2023-07 (h) USTC-2023-08

(i) USTC-2023-09 (j) USTC-2023-10 (k) USTC-2023-11 (l) USTC-2023-12

(m) USTC-2023-13 (n) USTC-2023-14 (o) USTC-2023-15 (p) USTC-2023-16

(q) USTC-2023-17 (r) USTC-2023-18 (s) USTC-2023-19 (t) USTC-2023-20

Fig. 8. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2023 under *VMAF (MSE model)* metric.



(a) USTC-2022-01 (b) USTC-2022-02 (c) USTC-2022-03 (d) USTC-2022-04

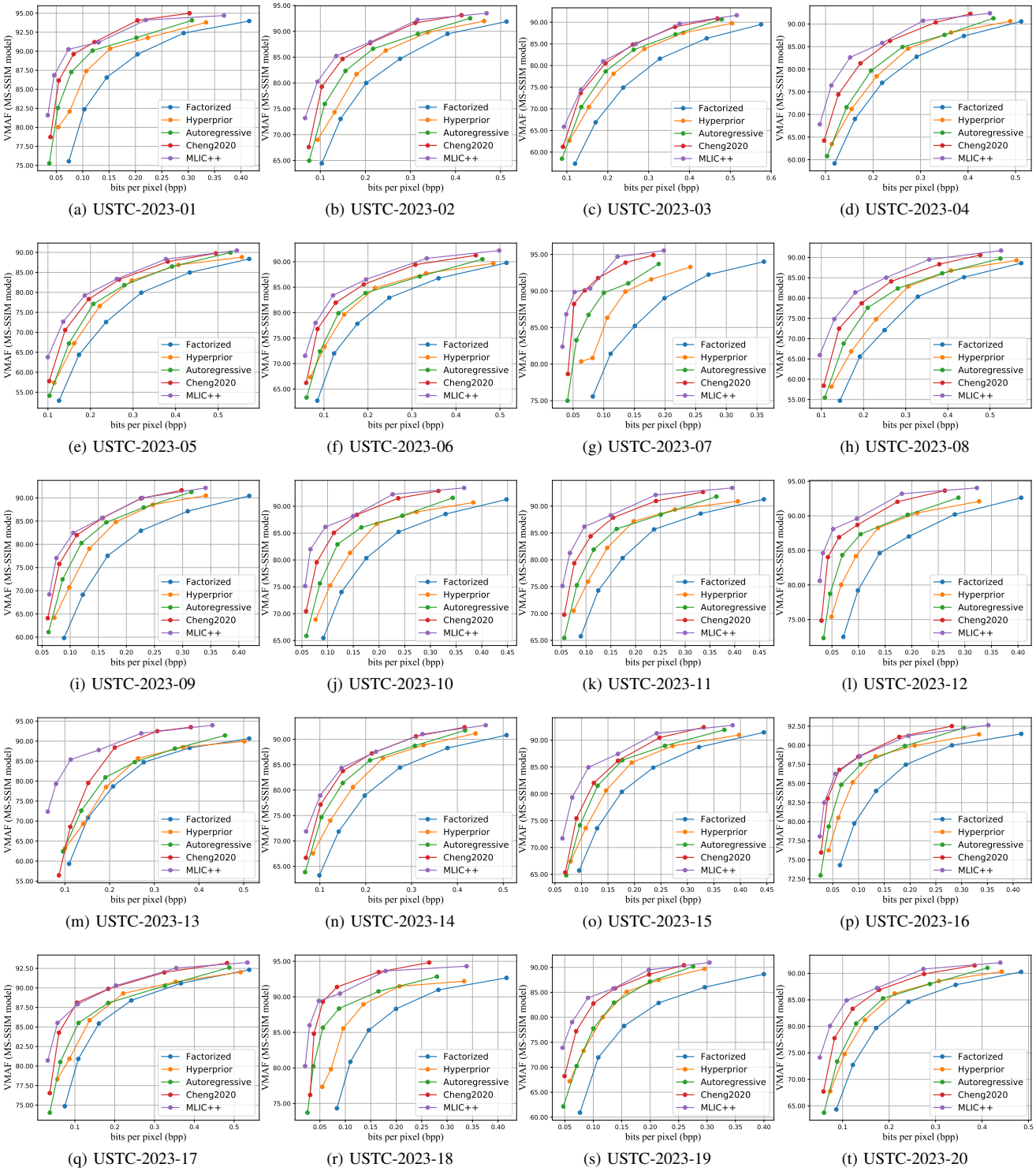
(e) USTC-2022-05 (f) USTC-2022-06 (g) USTC-2022-07 (h) USTC-2022-08

(i) USTC-2022-09 (j) USTC-2022-10 (k) USTC-2022-11 (l) USTC-2022-12

(m) USTC-2022-13 (n) USTC-2022-14 (o) USTC-2022-15 (p) USTC-2022-16

(q) USTC-2022-17 (r) USTC-2022-18 (s) USTC-2022-19 (t) USTC-2022-20

Fig. 9. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2022 under *VMAF (MS-SSIM model)* metric.



(a) USTC-2023-01 (b) USTC-2023-02 (c) USTC-2023-03 (d) USTC-2023-04

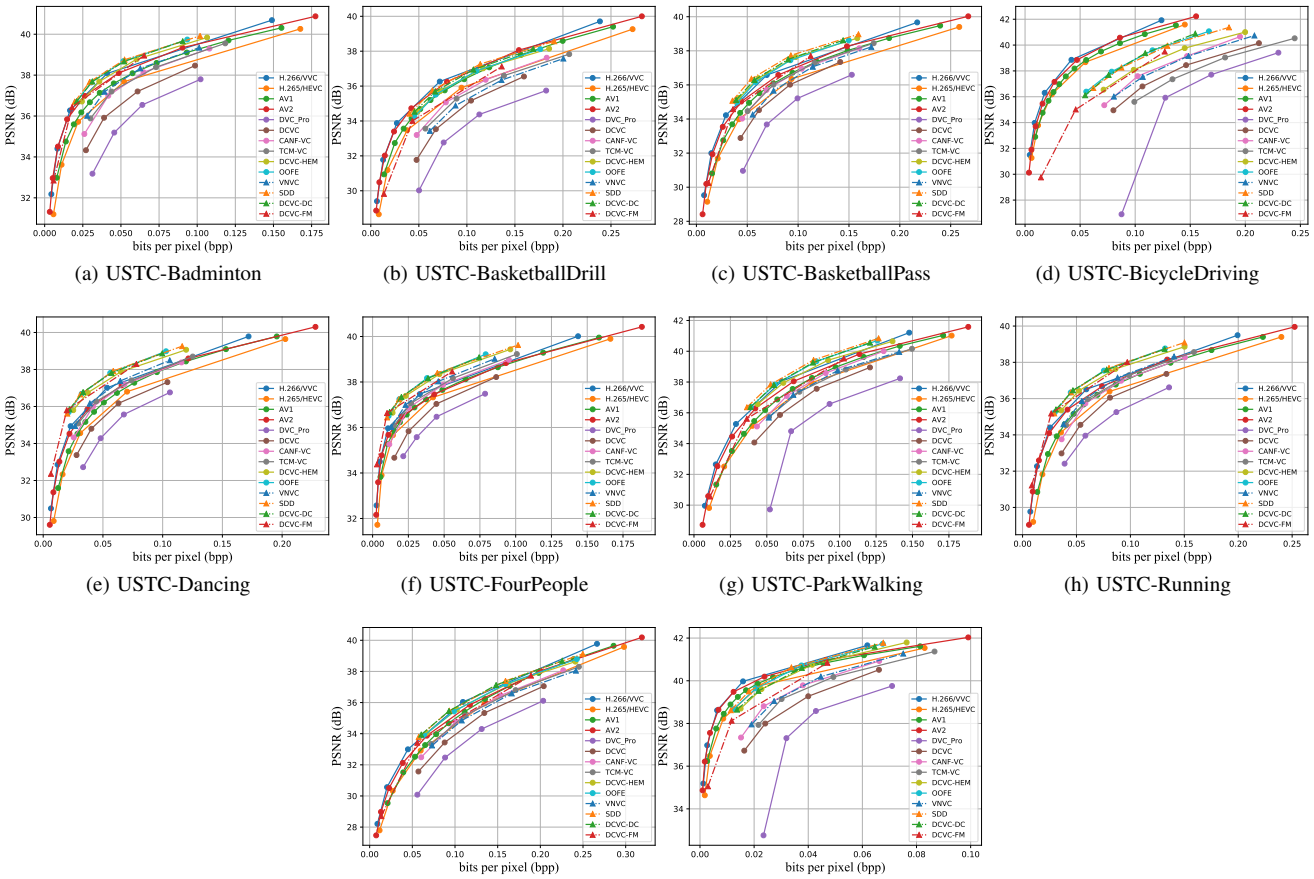
(e) USTC-2023-05 (f) USTC-2023-06 (g) USTC-2023-07 (h) USTC-2023-08

(i) USTC-2023-09 (j) USTC-2023-10 (k) USTC-2023-11 (l) USTC-2023-12

(m) USTC-2023-13 (n) USTC-2023-14 (o) USTC-2023-15 (p) USTC-2023-16

(q) USTC-2023-17 (r) USTC-2023-18 (s) USTC-2023-19 (t) USTC-2023-20

Fig. 10. Specific rate-distortion (RD) curves of advanced image compression schemes on each evaluative image of USTC-TD 2023 under *VMAF (MS-SSIM model)* metric.

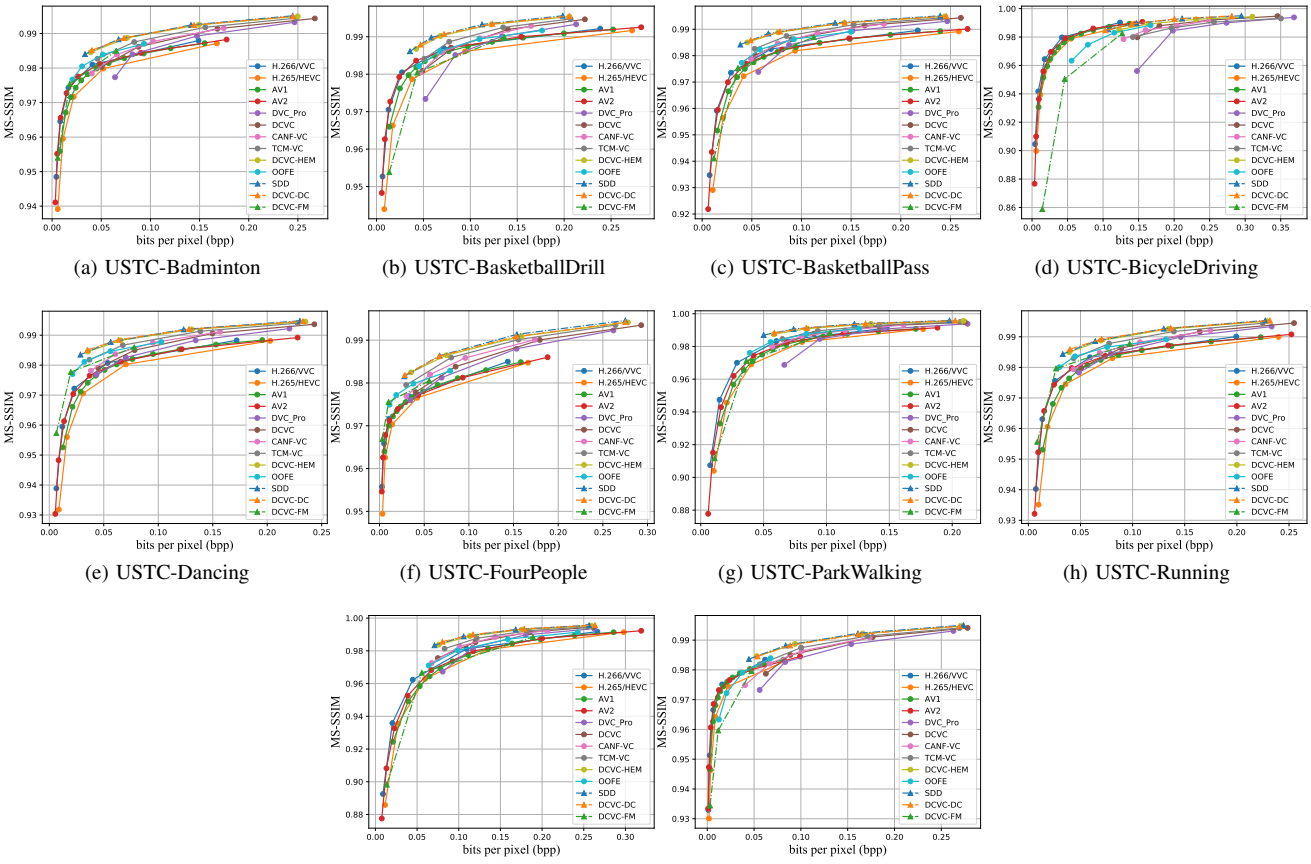


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 11. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD short (96 frames) video dataset under *PSNR* metric. The intra period is set to 32.

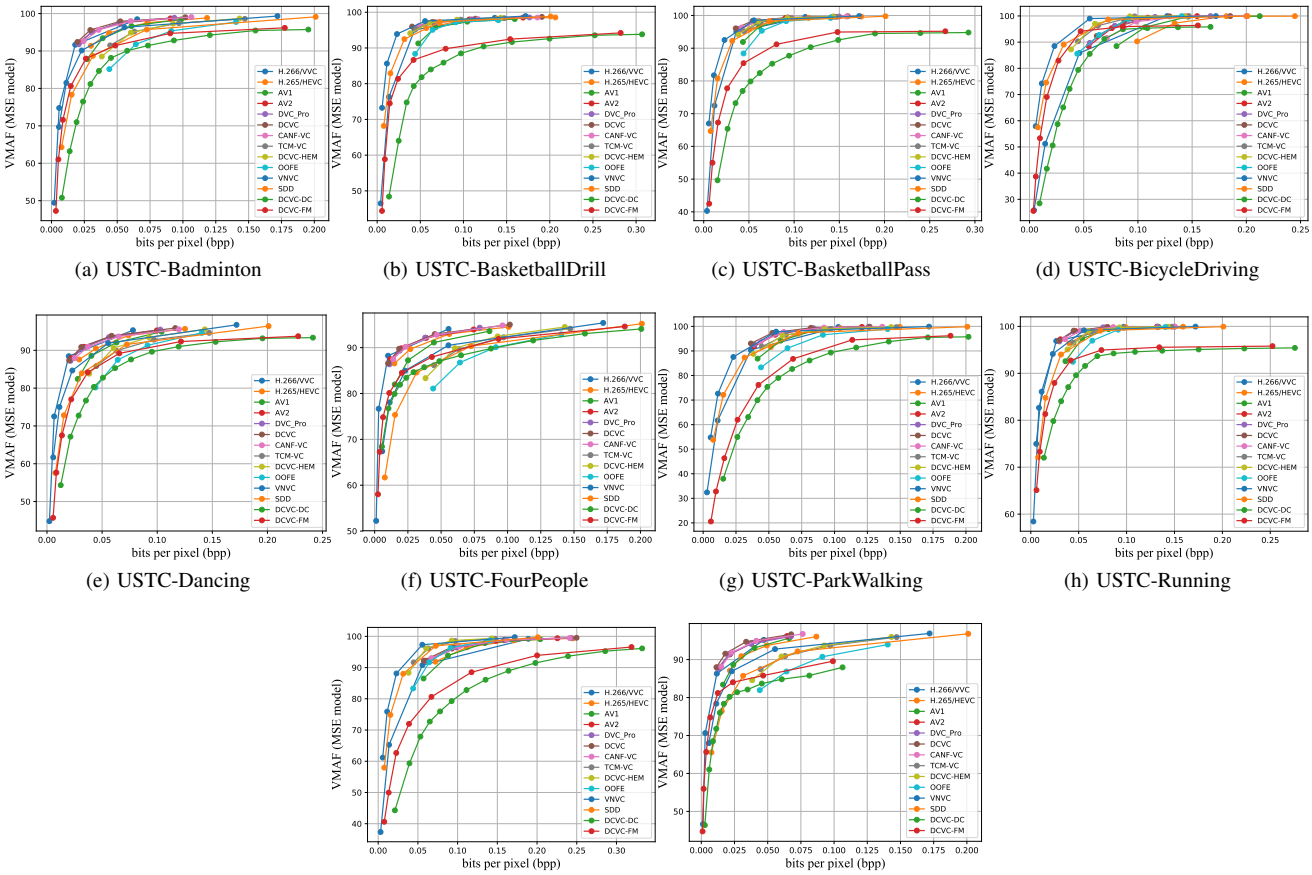


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 12. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD short (96 frames) video dataset under *MS-SSIM* metric. The intra period is set to 32.

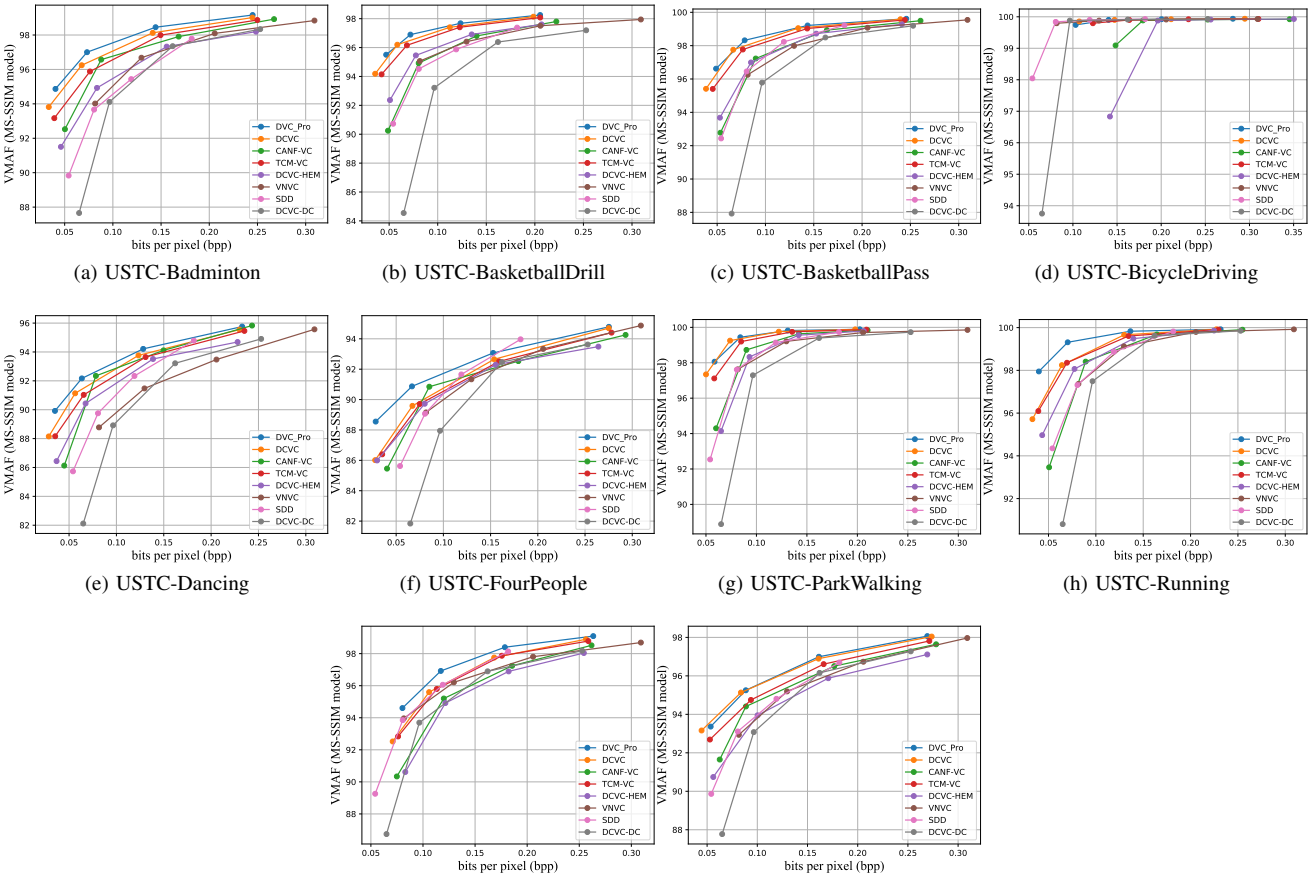


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 13. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD short (96 frames) video dataset under *VMAF (MSE model)* metric. The intra period is set to 32.

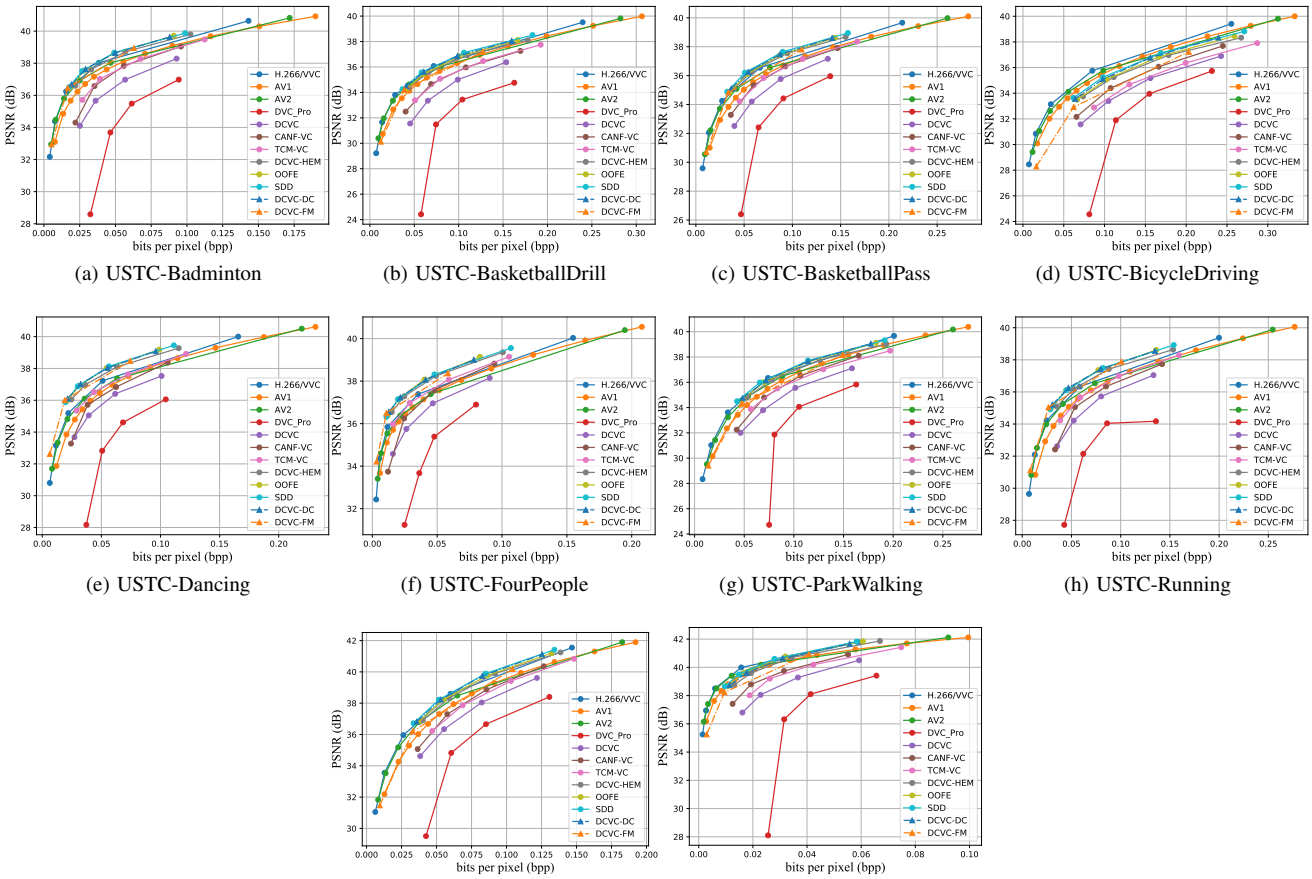


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 14. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD short (96 frames) video dataset under *VMAF (MS-SSIM model)* metric. The intra period is set to 32.

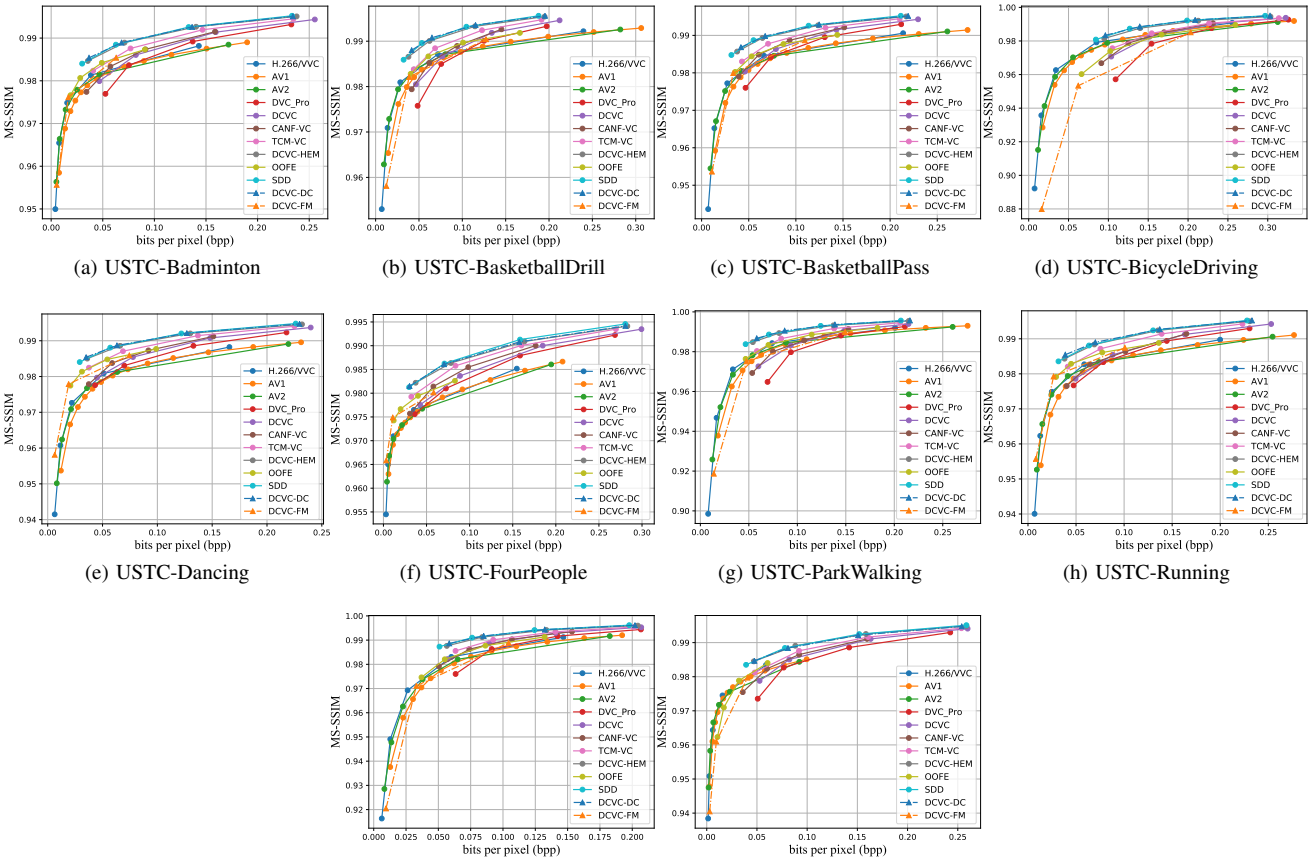


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 15. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *PSNR* metric. The intra period is set to 32.

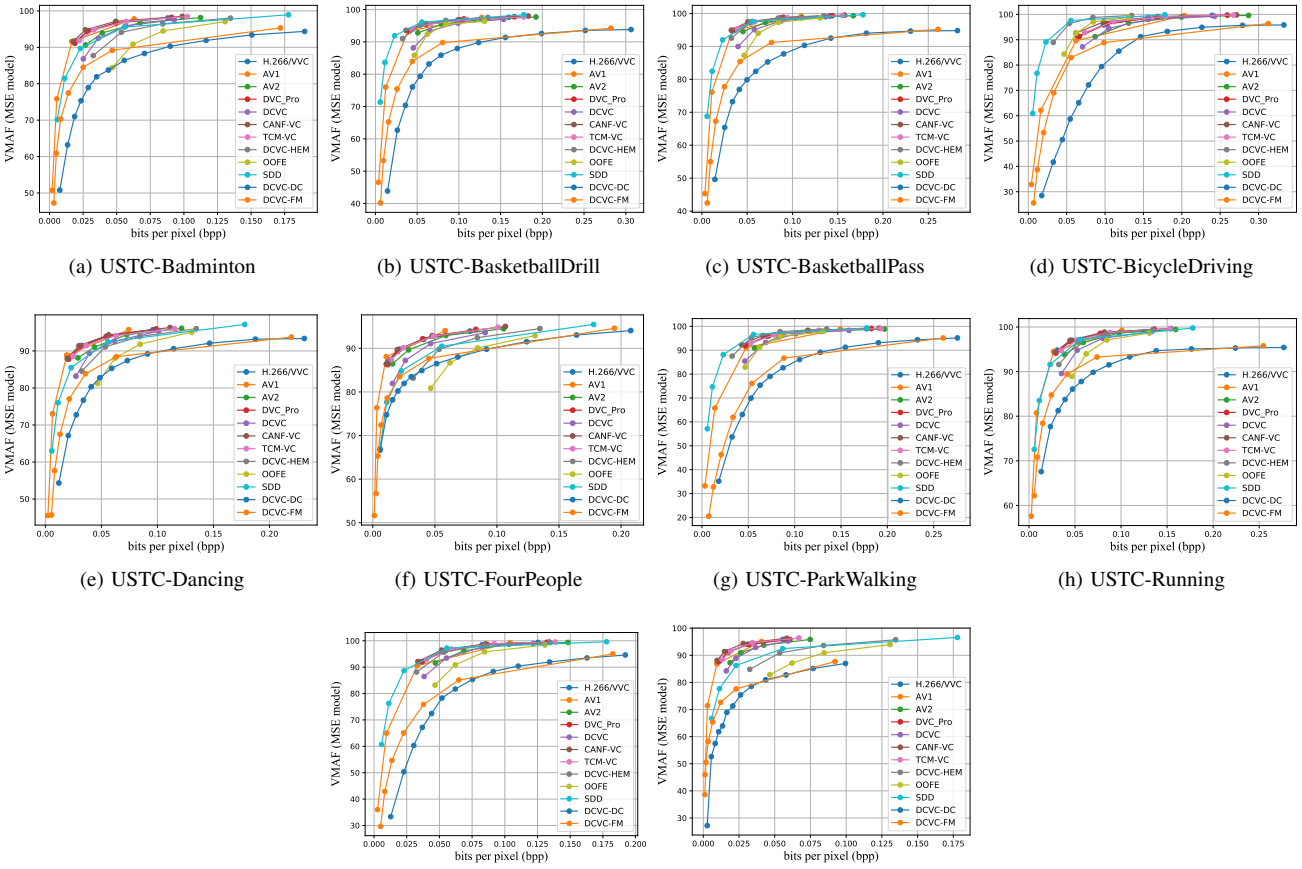


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 16. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *MS-SSIM* metric. The intra period is set to 32.

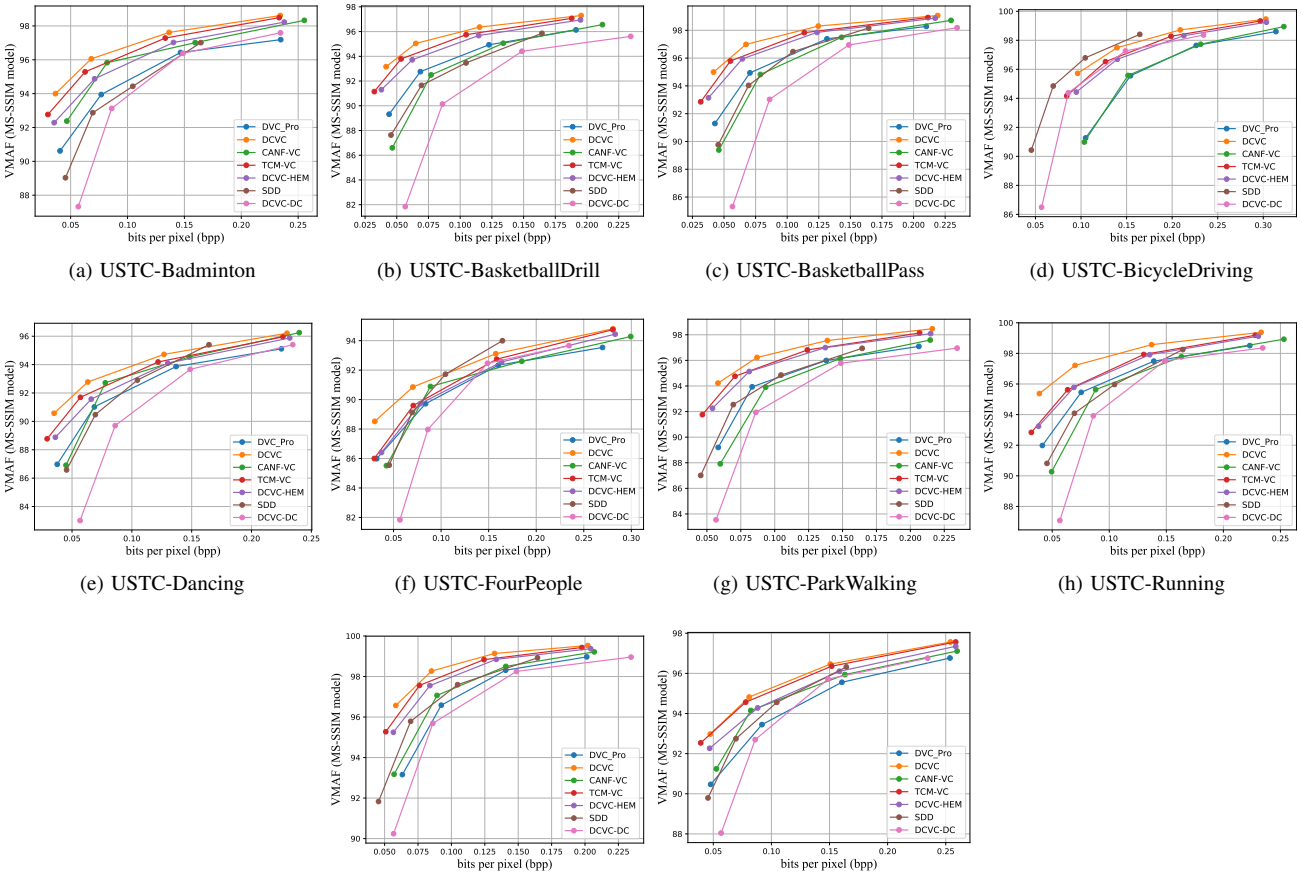


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 17. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *VMAF (MSE model)* metric. The intra period is set to 32.

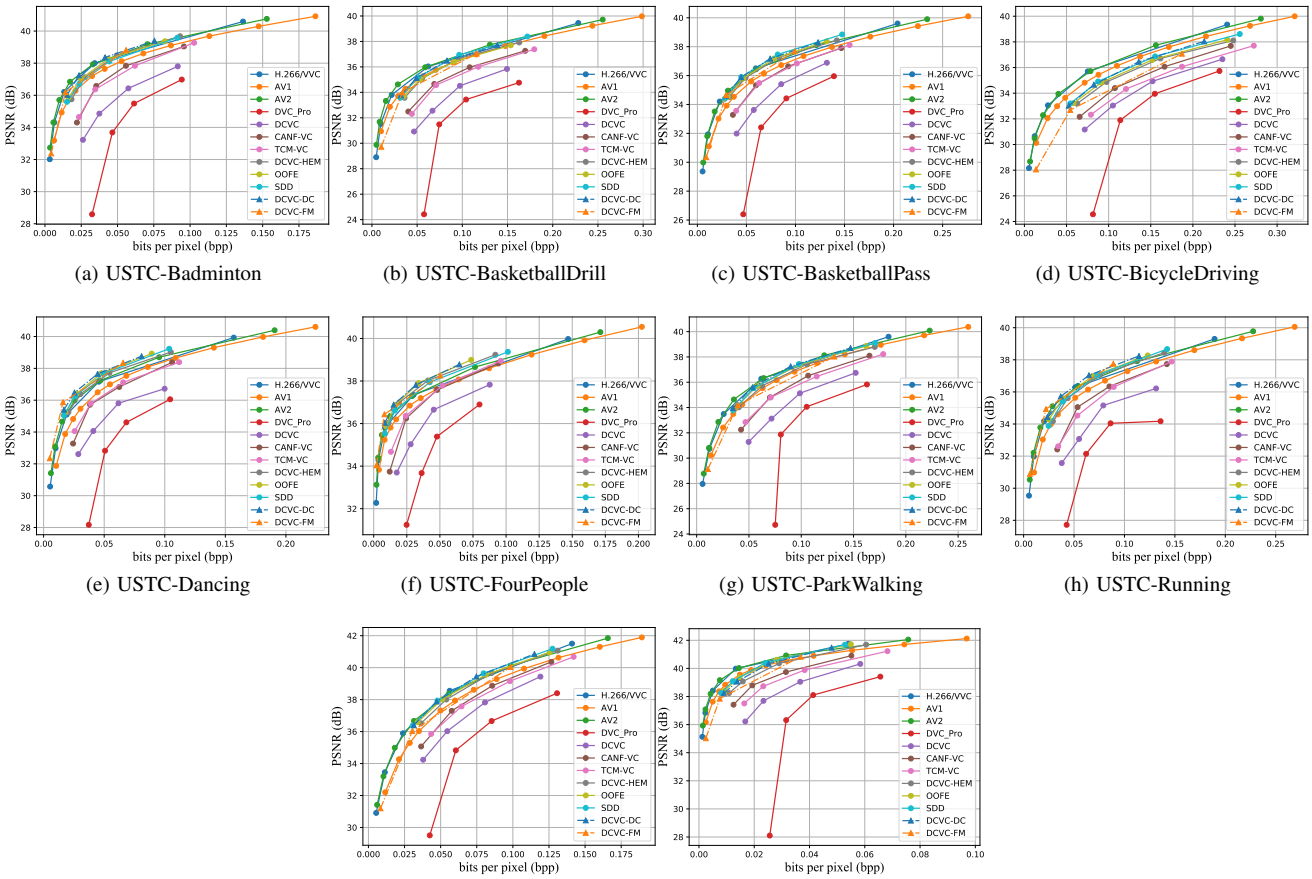


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 18. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *VMAF (MS-SSIM model)* metric. The intra period is set to 32.

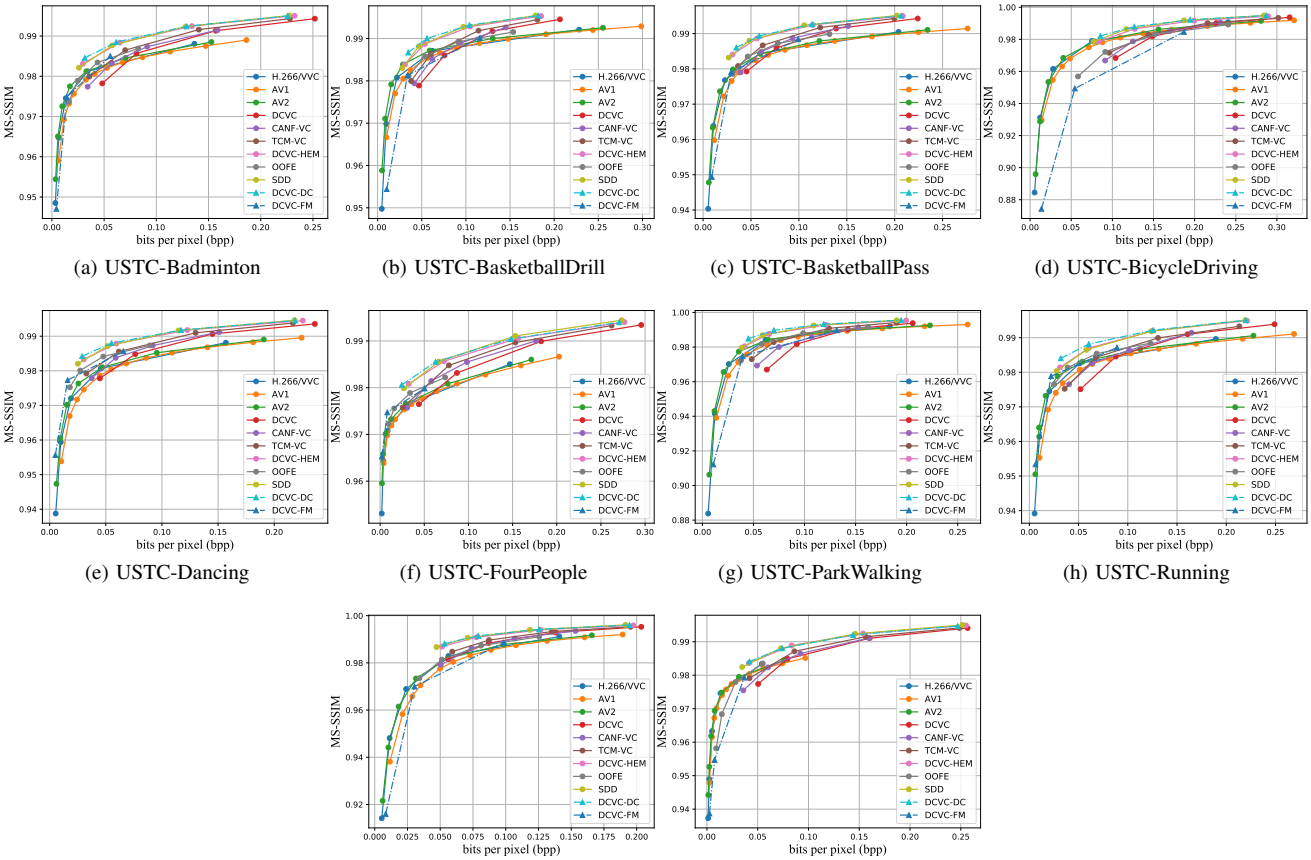


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 19. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *PSNR* metric. The intra period is set to -1.

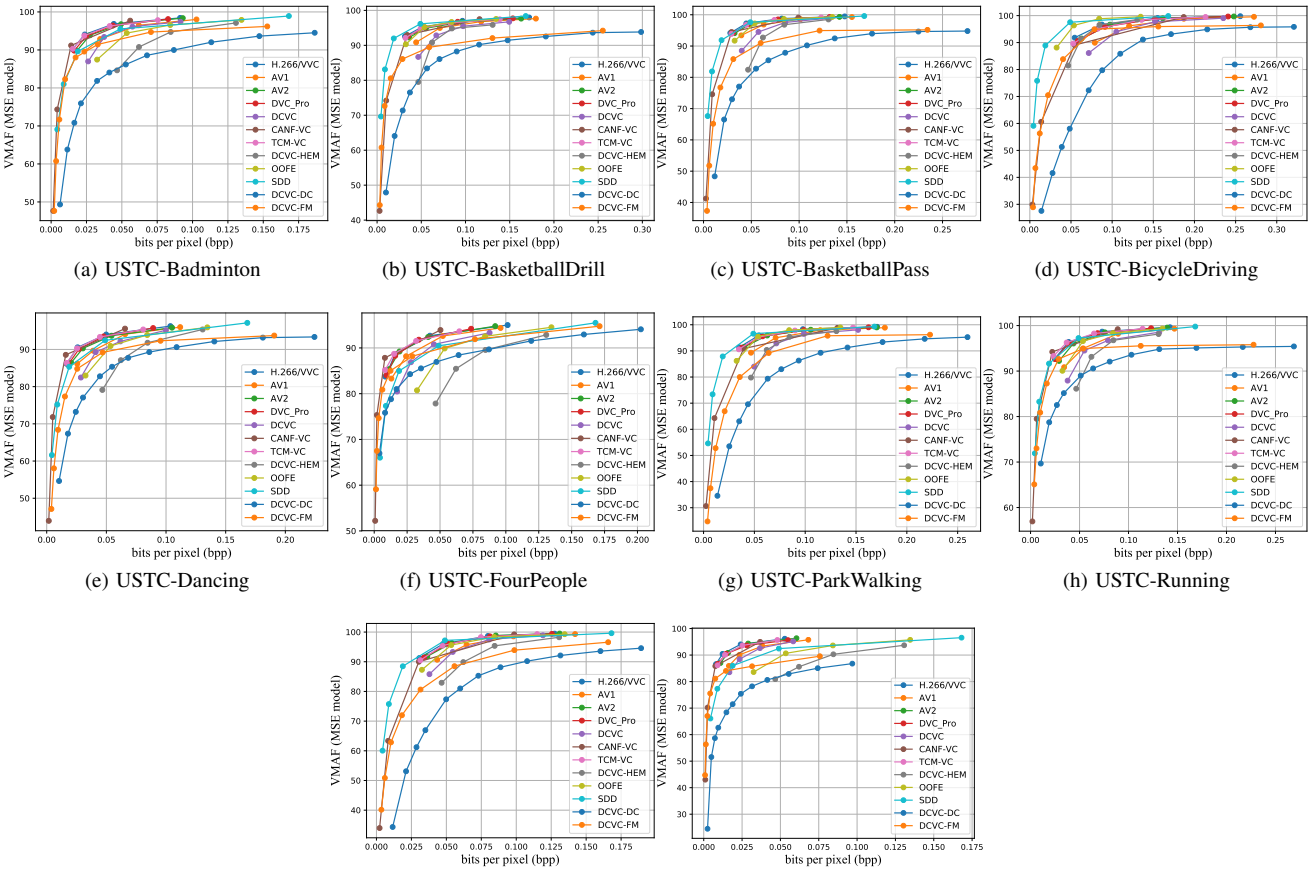


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 20. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *MS-SSIM* metric. The intra period is set to -1.

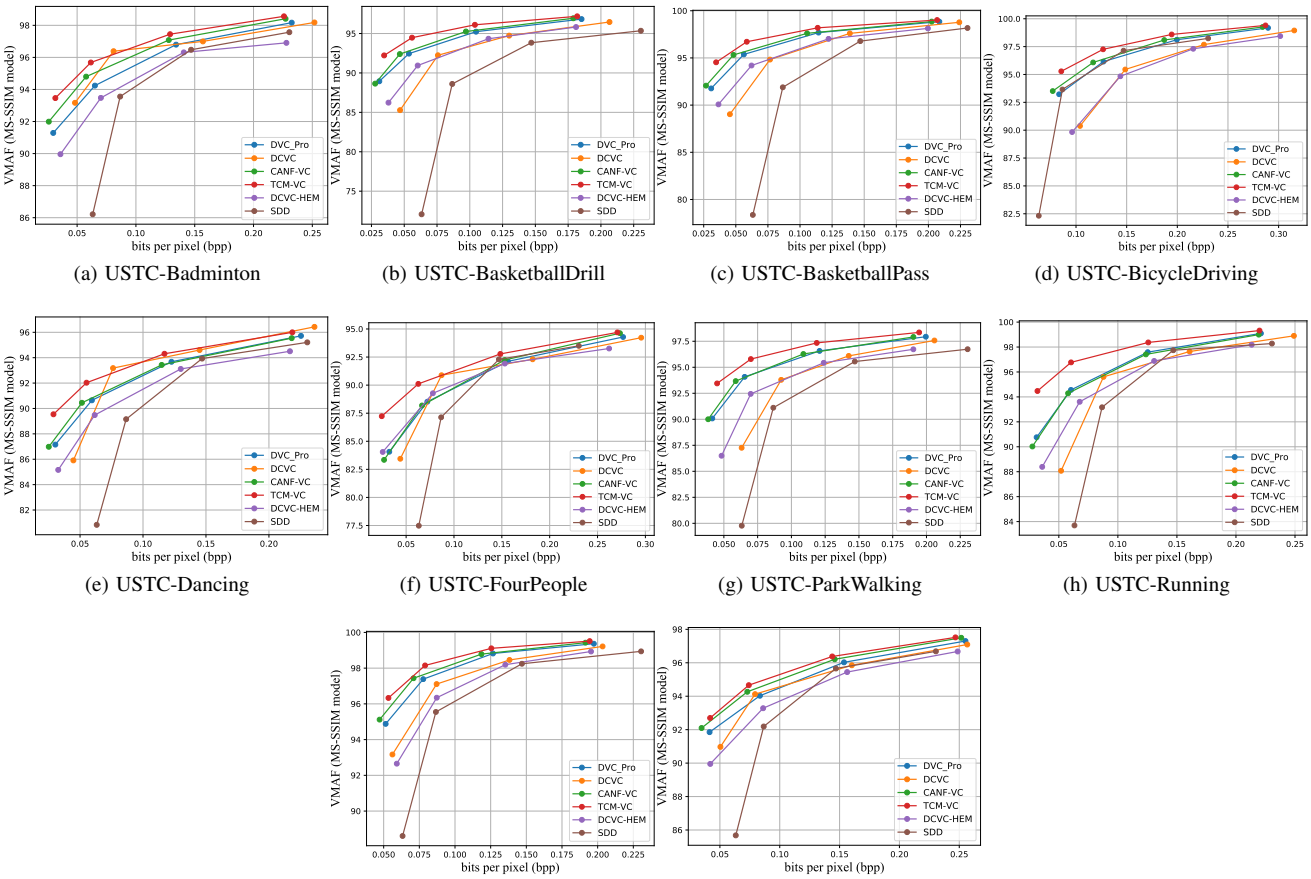


(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 21. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *VMAF (MSE model)* metric. The intra period is set to -1.



(a) USTC-Badminton (b) USTC-BasketballDrill (c) USTC-BasketballPass (d) USTC-BicycleDriving

(e) USTC-Dancing (f) USTC-FourPeople (g) USTC-ParkWalking (h) USTC-Running

(i) USTC-ShakingHands (j) USTC-Snooker

Fig. 22. Specific rate-distortion (RD) curves of advanced video compression schemes on each evaluative video of USTC-TD long video dataset under *VMAF (MS-SSIM model)* metric. The intra period is set to -1.

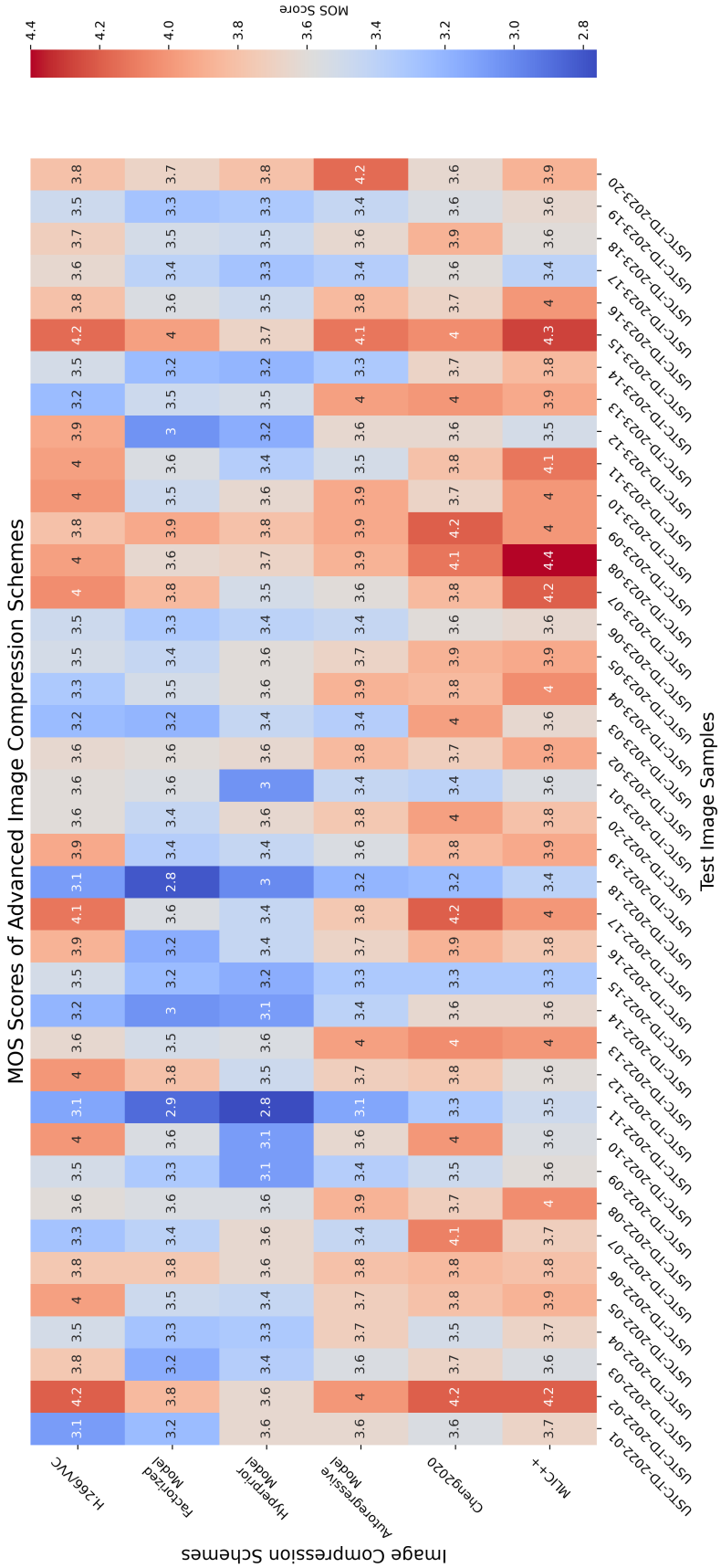


Fig. 23. Detailed *MOS* results of compressed images of different image compression schemes over USTC-TD 2022 and 2023 image datasets. The horizontal axis represents the abbreviation of specific test images, the vertical axis represents the abbreviation of different image compression test schemes, and the different colors of the heatmap indicate the distribution of different *MOS* scores.

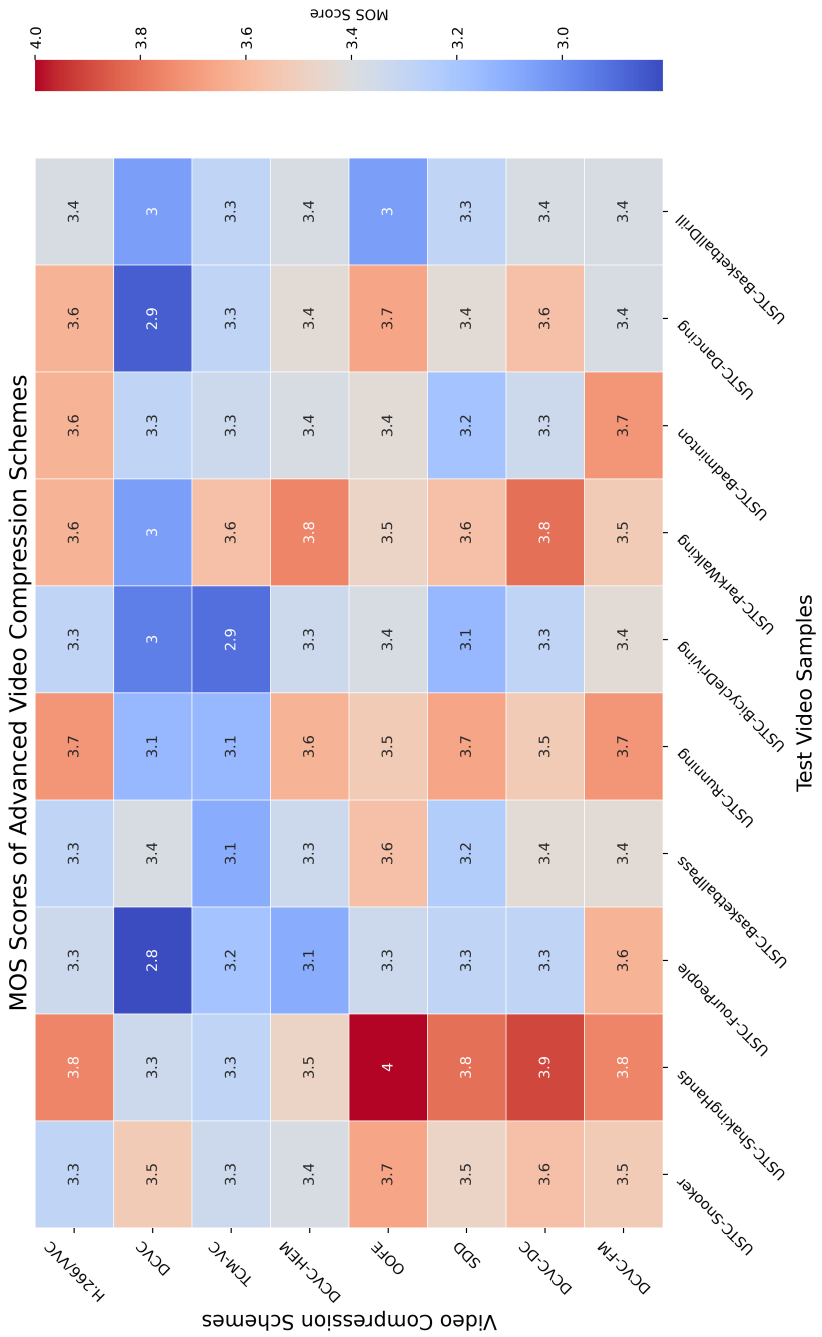


Fig. 24. Detailed *MOS* results of compressed videos of different video compression schemes over USTC-TD 2023 video dataset. The horizontal axis represents the abbreviation of specific test videos, the vertical axis represents the abbreviation of different video compression test schemes, and the different colors of the heatmap indicate the distribution of different *MOS* scores.