Scale-space flow for end-to-end optimized video compression
———— Supplementary Materials ————

Anonymous CVPR submission

Paper ID 9692

Contents

1. Baselines

Following prior works, we used *ffmpeg*[1] to produce the evaluation metrics for H.264 and HEVC. Below are the exact commands we used:

H.264 (medium)

```
ffmpeg -i FILE.y4m \
    -c:v h264 -crf CRF \
    -preset medium -bf 0 FILE.mp4
```

where 'FILE' represents the input file and 'CRF' the quality level.

HEVC (medium)

```
ffmpeg -i FILE.y4m \
    -c:v hevc -crf CRF  \
    -preset medium \
    -x265-params bframes=0 \
    FILE.mp4
```

Encoder	Decoder
Conv, 5x5/2, 128ch	DeConv, 5x5/2, 128ch
ReLu	ReLu
Conv, 5x5/2, 128ch	DeConv, 5x5/2, 128ch
ReLu	ReLu
Conv, 5x5/2, 128ch	DeConv, 5x5/2, 128ch
ReLu	ReLu
Conv, 5x5/2, 192ch	DeConv, 5x5/2, 3ch

Table 1. The Encoder and Decoder architecture used by the model. Each row denotes a (De)Conv layer or an activation function, where "KxK/s C ch" denotes C output channels and a kernel size of K and a stride of S.

Hyper Encoder	Hyper Decoder (scale)	Hyper Decoder (mean)
Conv, 5x5/2, 192ch	DeConv, 5x5/2, 192ch	DeConv, 5x5/2, 192ch
ReLu	QReLu	ReLu
Conv, 5x5/2, 192ch	DeConv, 5x5/2, 192ch	DeConv, 5x5/2, 192ch
ReLu	QReLu	ReLu
Conv, 5x5/2, 192ch	DeConv, 5x5/2, 192ch	DeConv, 5x5/2, 192ch
	QReLu	

Table 2. The Hyperprior architecture used by the model. Each row denotes a (De)Conv layer or an activation function, where "KxK/s, C ch" denotes C output channels and a kernel size of K and a stride of S.

where 'FILE' represents the input file and 'CRF' the quality level.

HEVC (very-fast) Here we used the same settings as reported in [4]:

```
ffmpeg -i FILE.y4m \
    -c:v hevc -crf CRF \
    -preset veryfast  \
    -x265-params bframes=0 \
    -tune zerolatency \
    -x265-params "keyint=12:verbose=1" \
    FILE.mp4
```

where 'FILE' represents the input file and 'CRF' the quality level.

Metric computation For computing frame-level metrics (such as PSNR and MS-SSIM), we dumped the original/encoded videos to frames to sRGB as follows

```
    ffmpeg -i FILE  FILE_%05d.png
```

We also used this command to obtain the frames that we feed to our models during evaluation.

2. Architecture

In Tables 1&2 we show the full Encoder/Decoder and Hyperprior architecture[**?**] used in our model. We repurposed this architecture for the 'Image', 'Scale Space Flow' and 'Residual' branches depicted in Figure 2 in the main paper. Following [2] we used QReLu for the Hyperprior to enable deterministic decoding.

3. Calculating the Bjøntegaard Delta Bit Rate

The Bjøntegaard Delta bitrate (BD rate) summarizes the relative performance between two compression models by calculating the relative rate savings for equal quality averaged across the shared quality range. This value is expressed as a percentage [3]. For example, if a codec has a BD rate of 5% relative to H.264, we would expect videos with the same visual quality but which require 5% less space than an H.264 encoding.

Due to the typical response of RD curves, a direct calculation of the average percent change between two RD curves will weight higher bit rates more than lower bit rates. For this reason, the BD rate is calculated on a log scale for bit rate. The original formulation fit a cubic polynomial to exactly four rate points. The area to the left of each curve was then calculated via numerical integration. Following common practice, we use a generalized variant of BD rate that relies on piecewise cubic hermite interpolating polynomials (PCHIP) for interpolation, which supports a larger number of rate points. Integration uses the trapezoid rule after resampling each RD curve at a sufficiently fine granularity (typically 100 points).

4. Subjective Comparison

Ours (0.043 bpp)

HEVC (0.0798 bpp)

Figure 1. Comparison between the proposed method and HEVC on a frame from a video where we outperform HEVC. Our reconstruction is slightly sharper while the average bpp is significantly smaller. (Video ‘Beauty‘ in UVG).

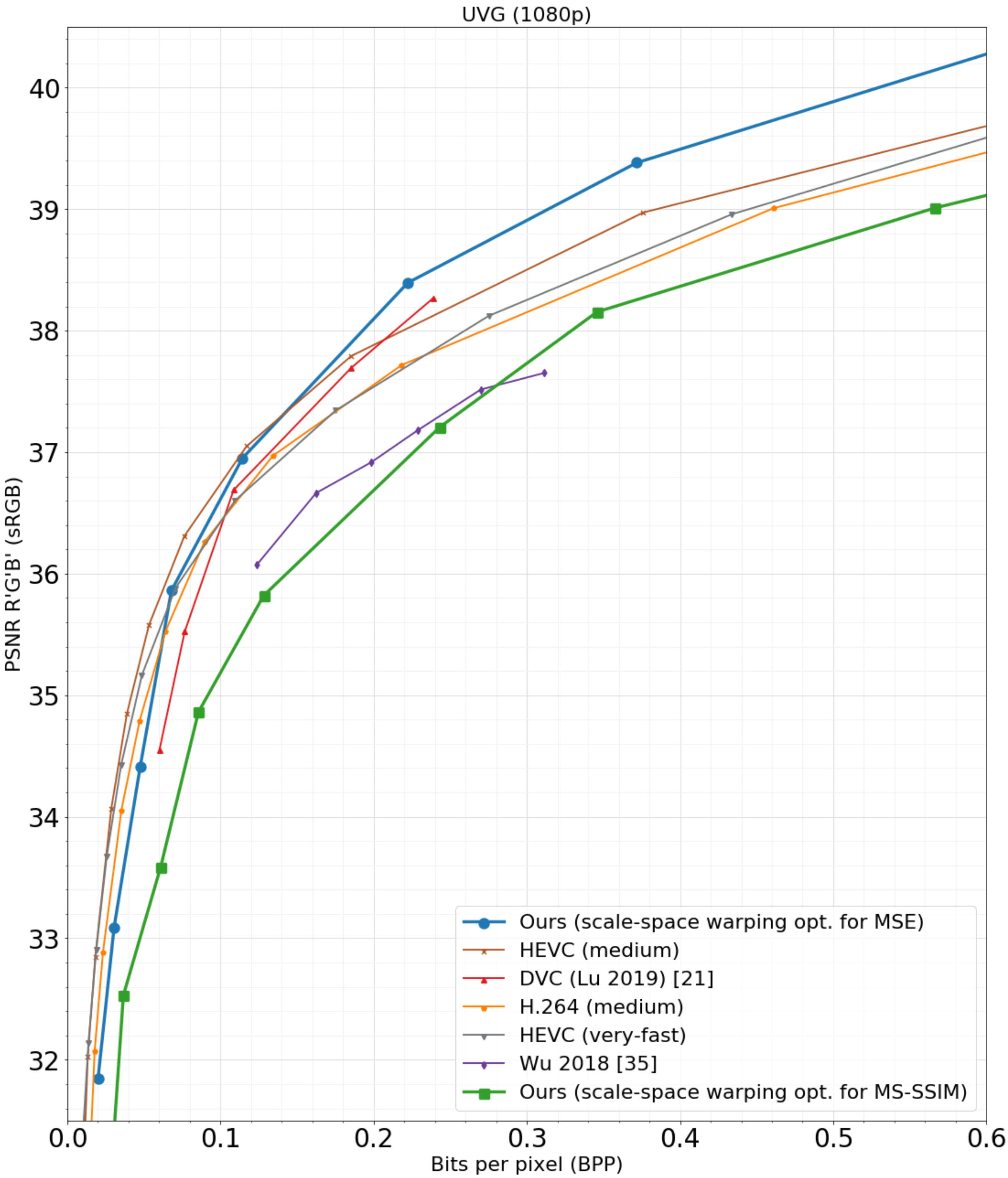
Ours (0.109 bpp)

HEVC (0.096 bpp)

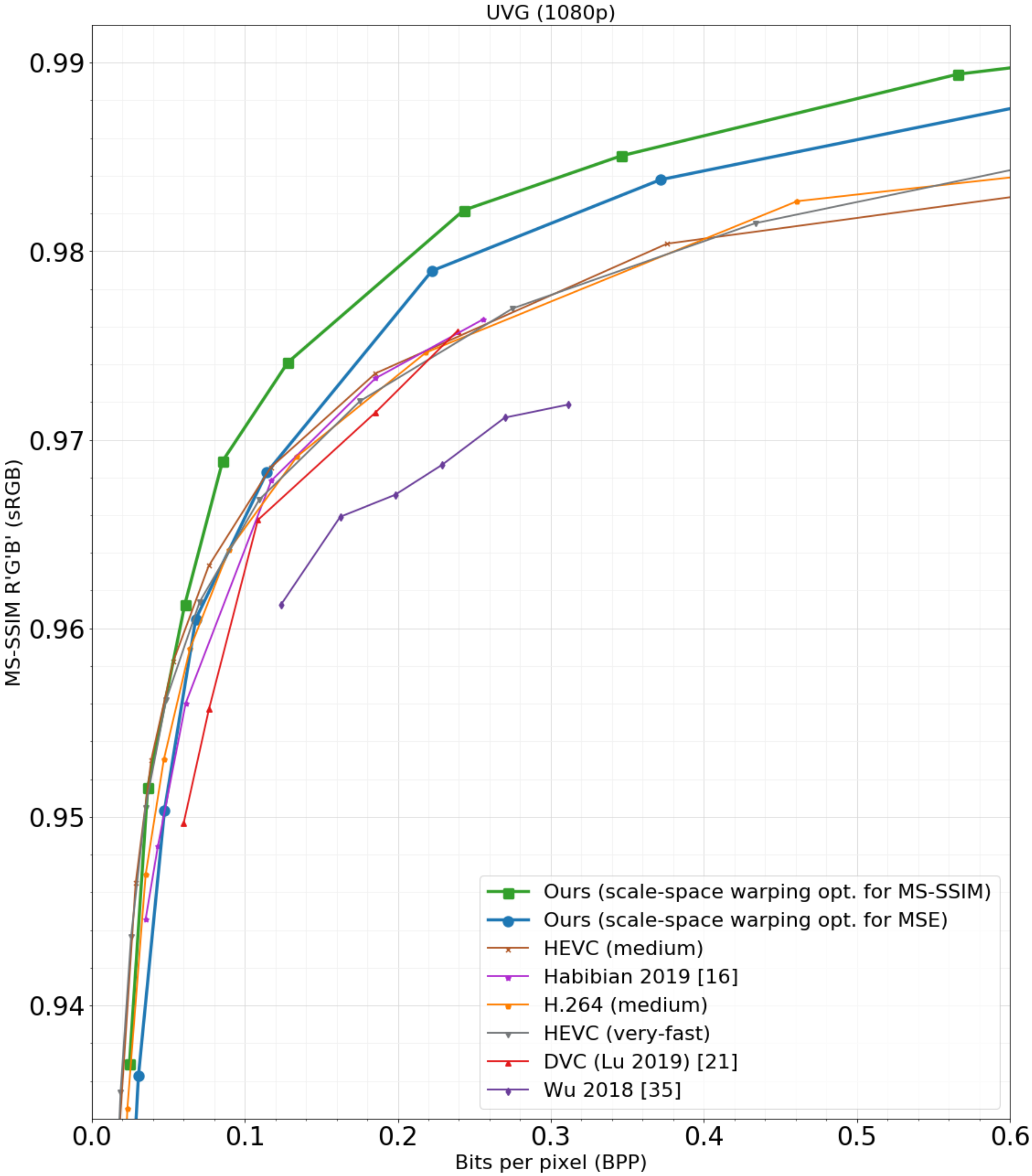
Figure 2. Comparison between the proposed method and HEVC on a frame from an animation, where our method performs poorly according to quantitative metrics. (Video 20 in MCL-JCV).

5. Aggregate Rate-Distortion Graphs

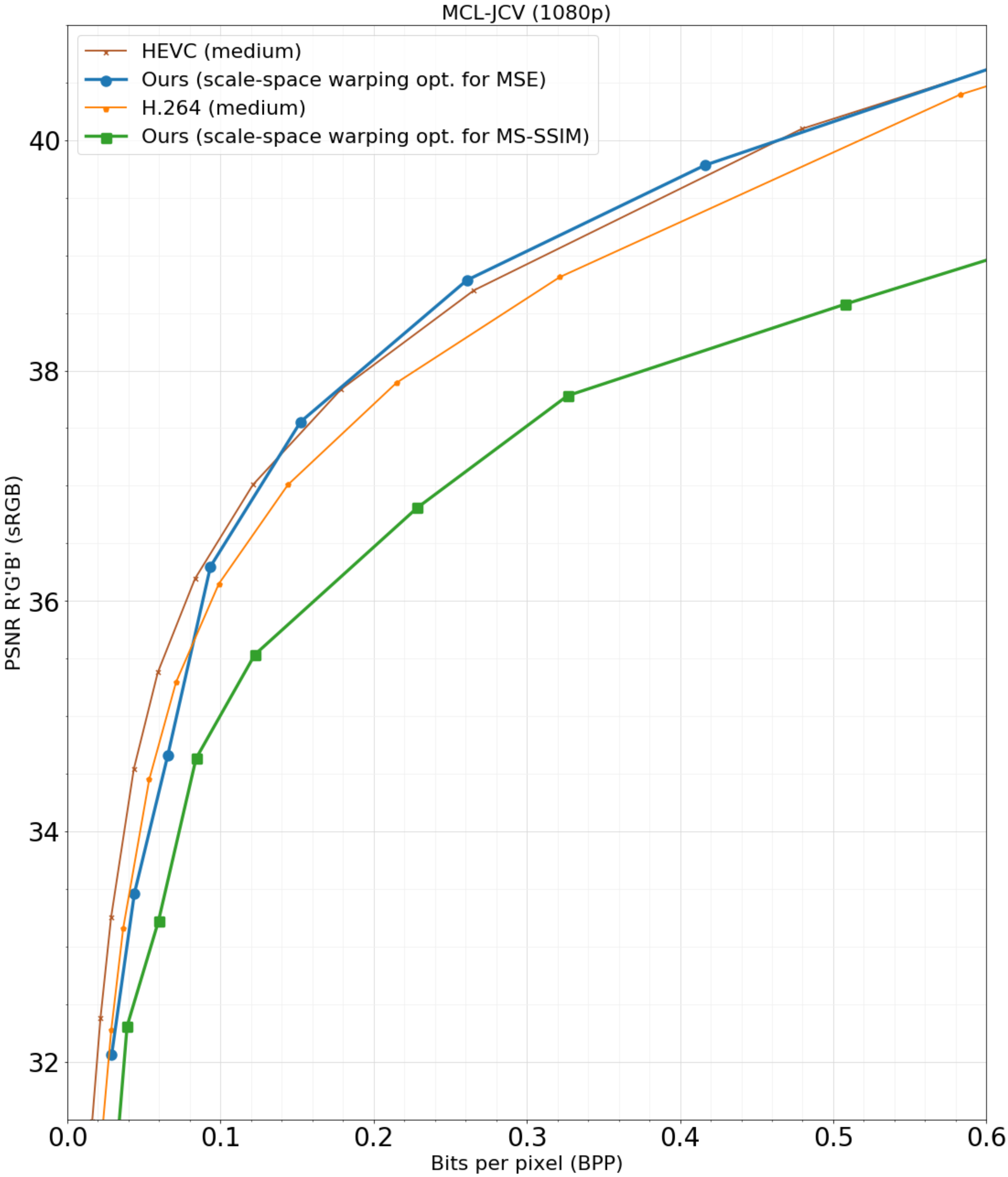
5.1. UVG (PSNR)



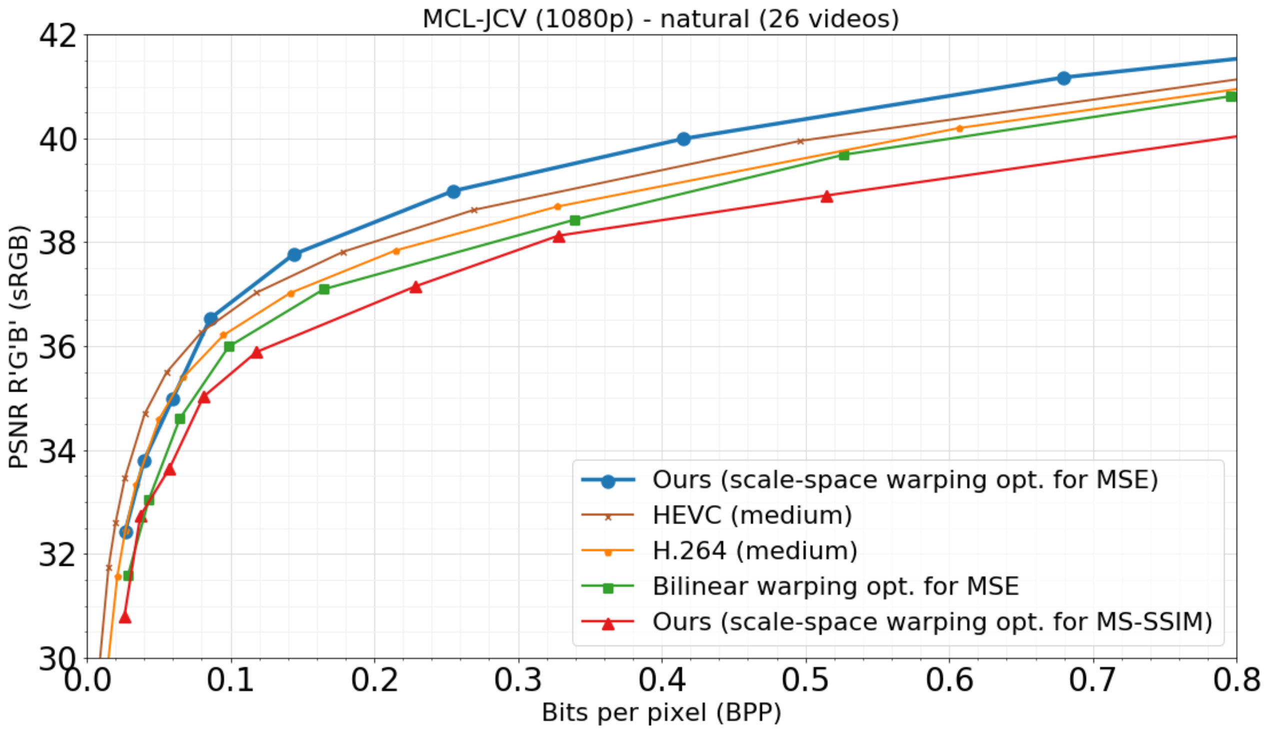
5.2. UVG (MS-SSIM)



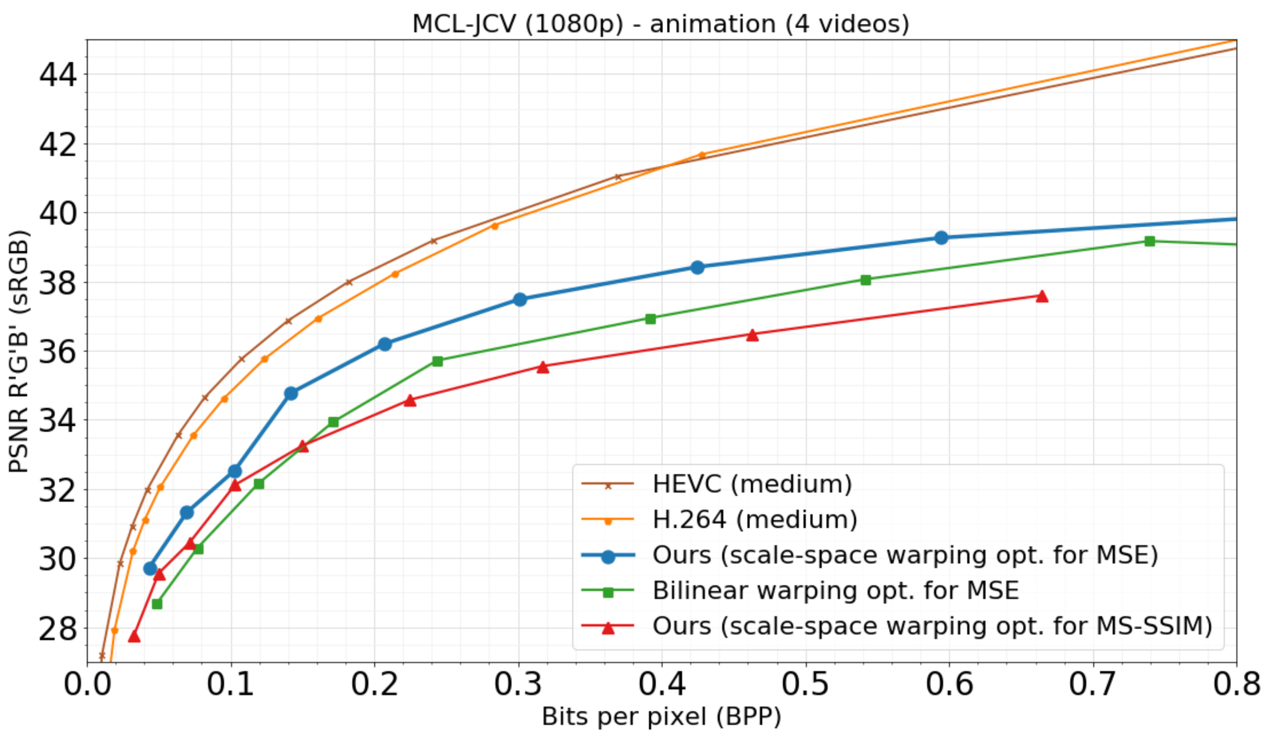
5.3. MCL-JCV (PSNR)



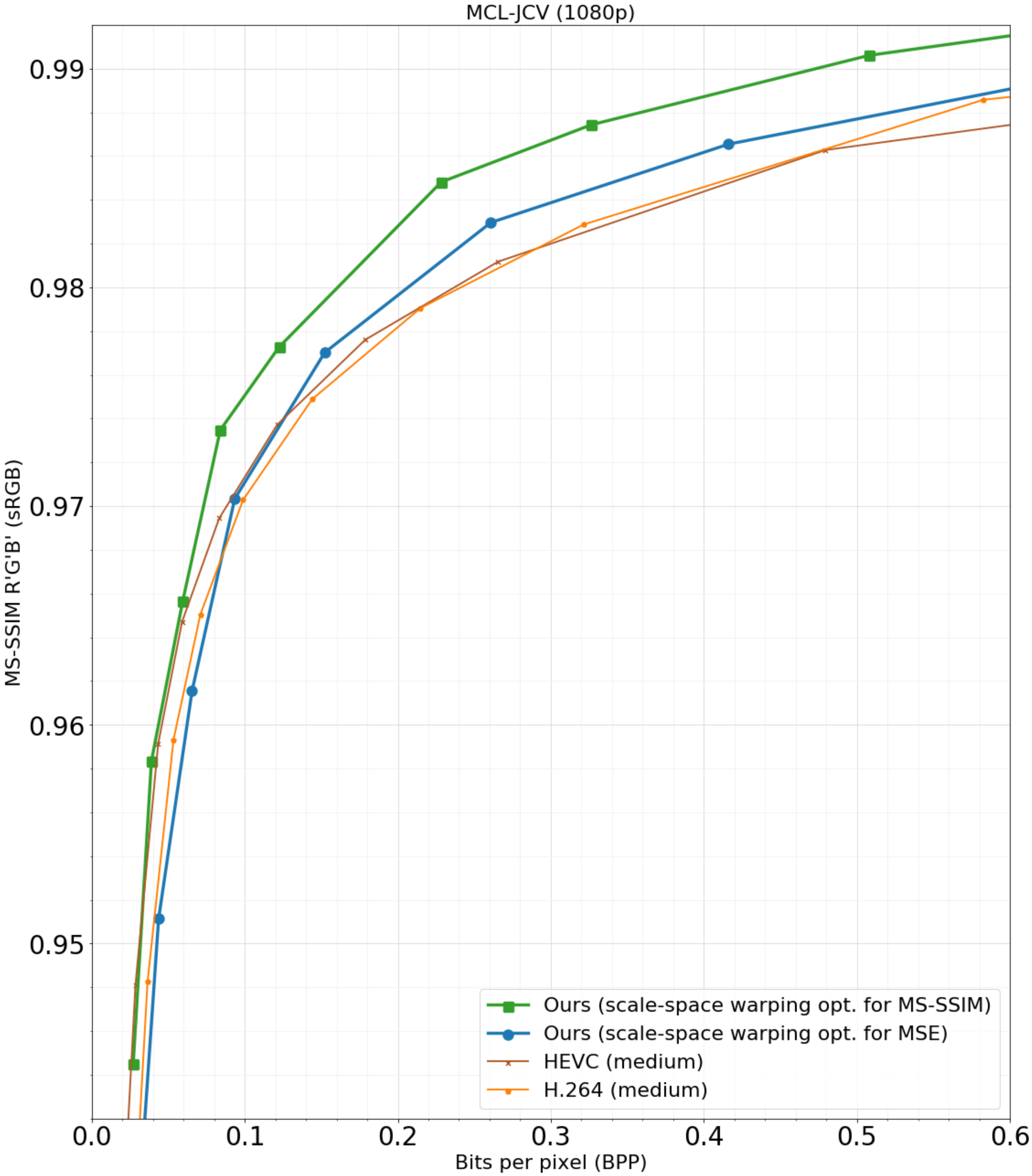
5.3.1 Only "Natural" Videos in MCL-JCV (PSNR)



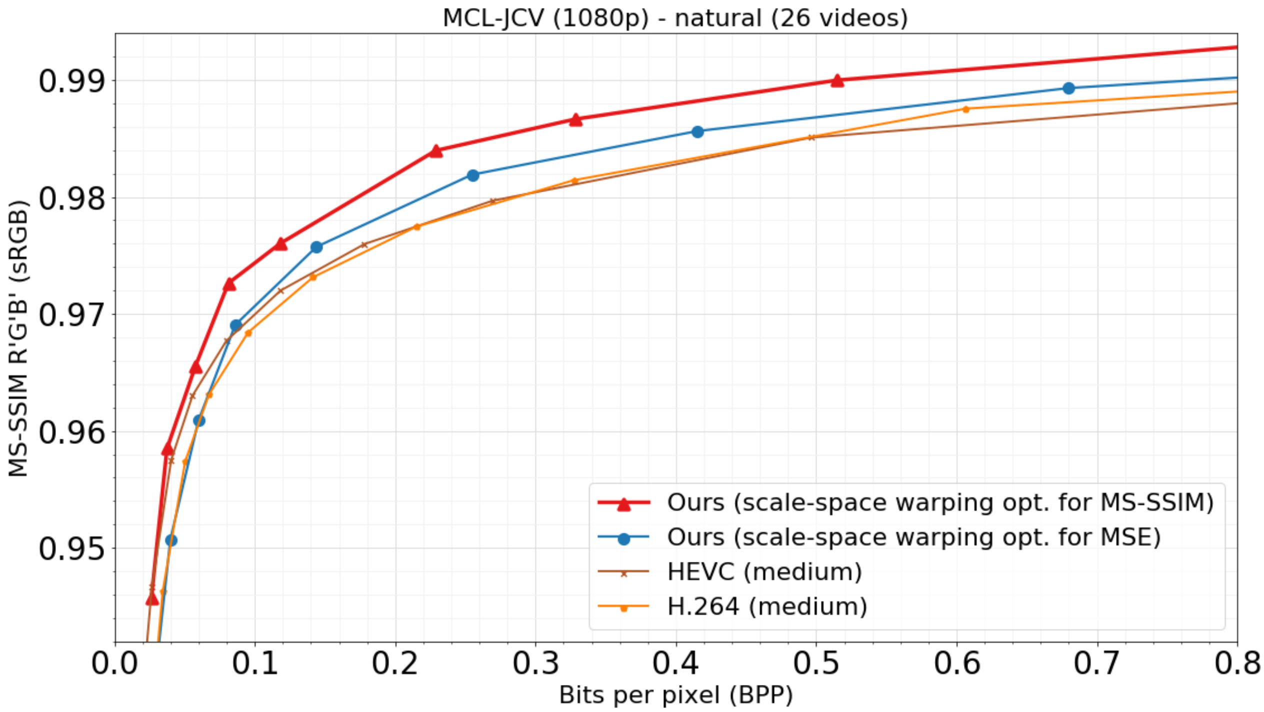
5.3.2 Only Animated Videos in MCL-JCV (PSNR)



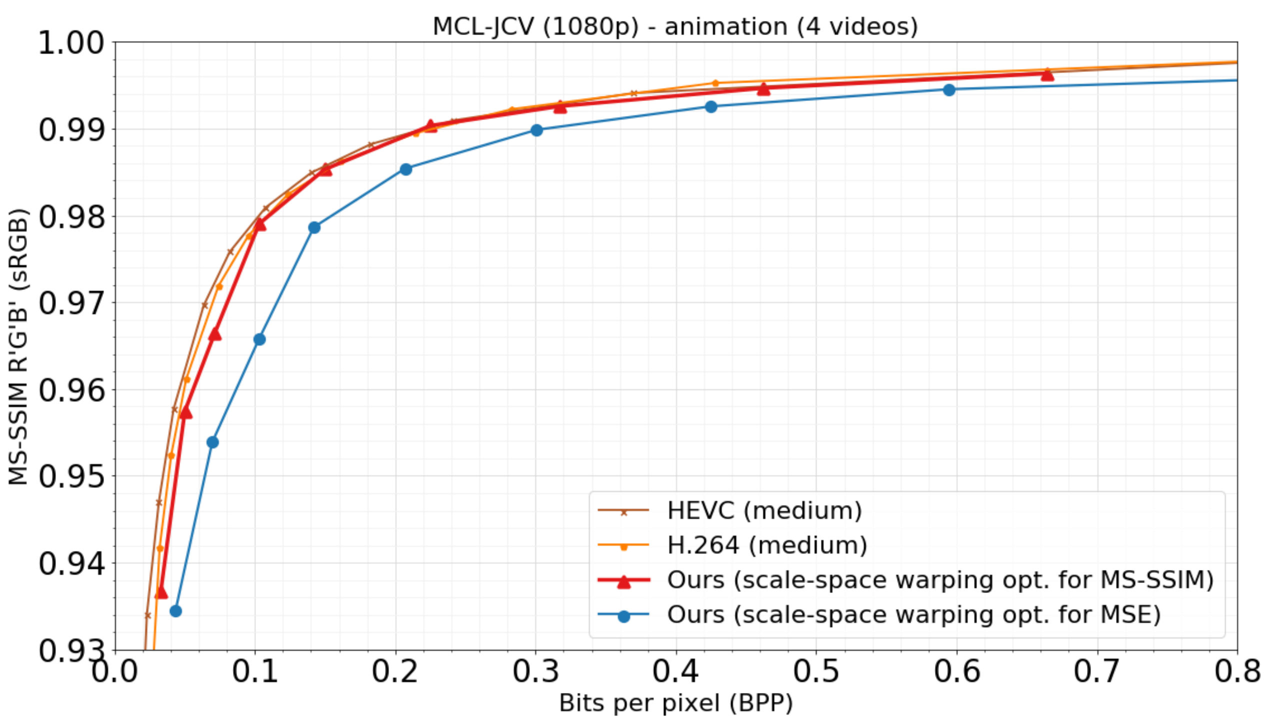
5.4. MCL-JCV (MS-SSIM)



5.4.1 Only "Natural" Videos in MCL-JCV (MS-SSIM)

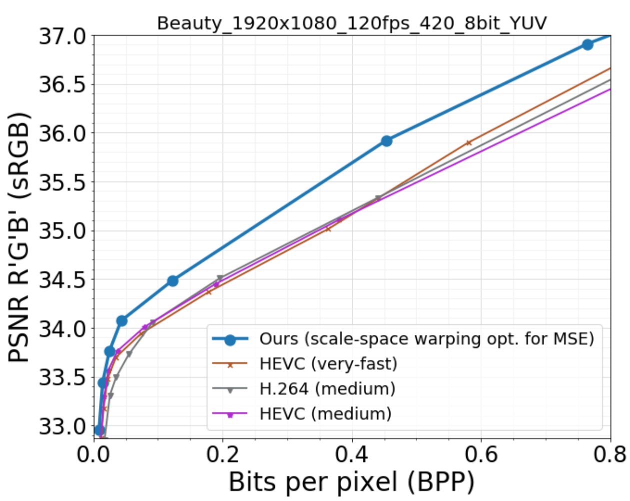


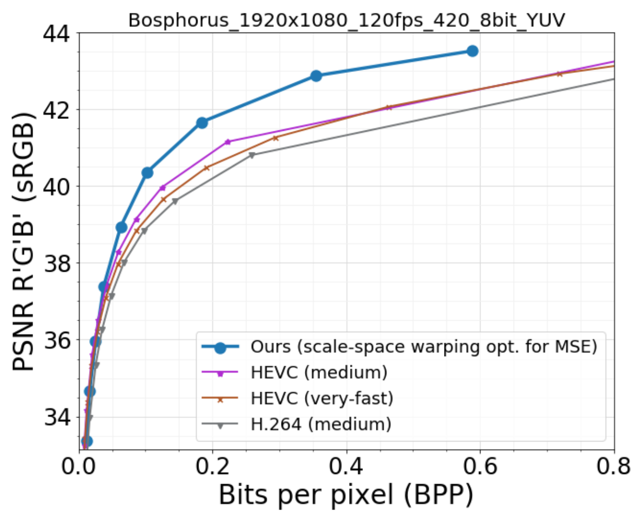
5.4.2 Only Animated Videos in MCL-JCV (MS-SSIM)

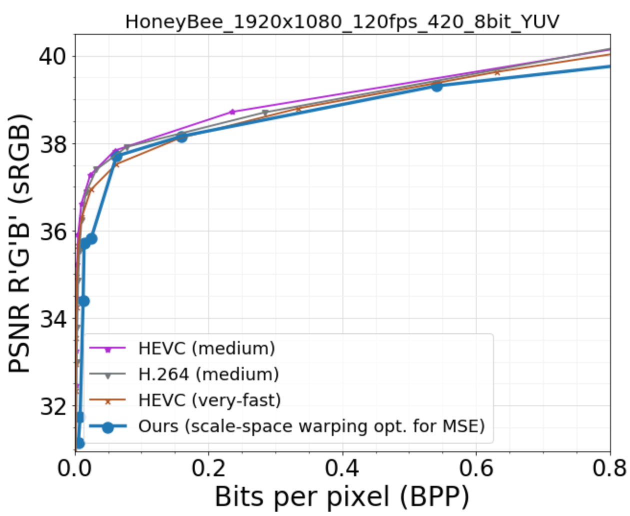


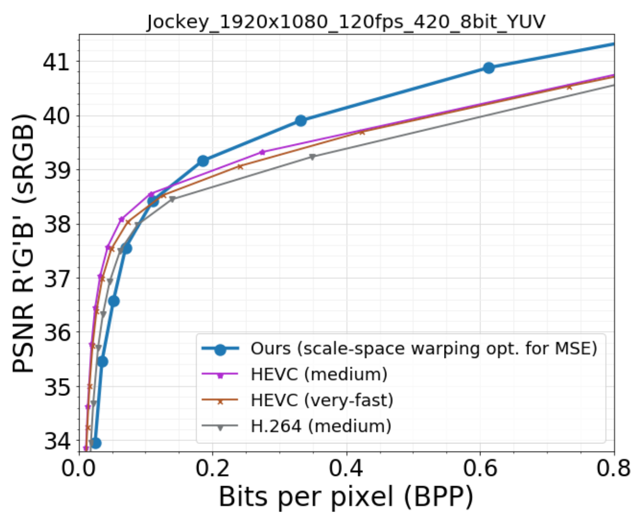
6. Per-Video Rate-Distortion Graphs

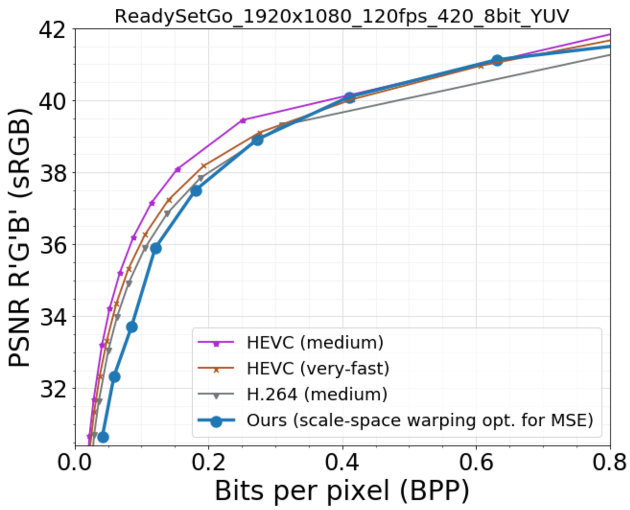
6.1. UVG

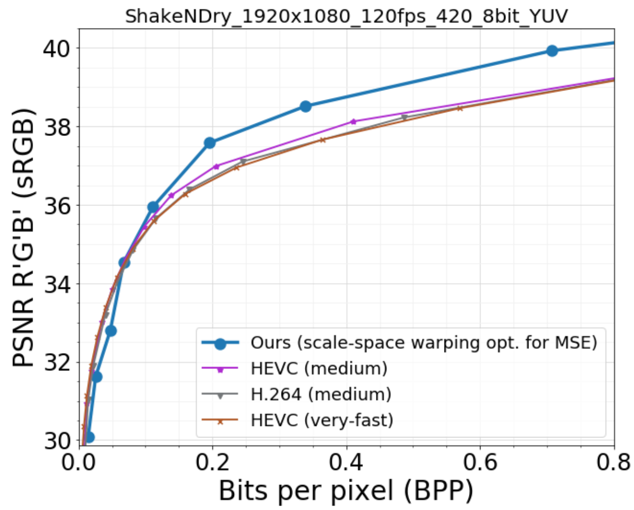


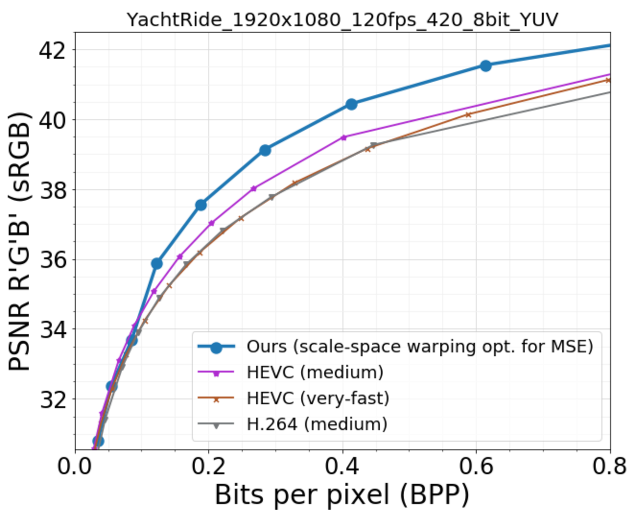




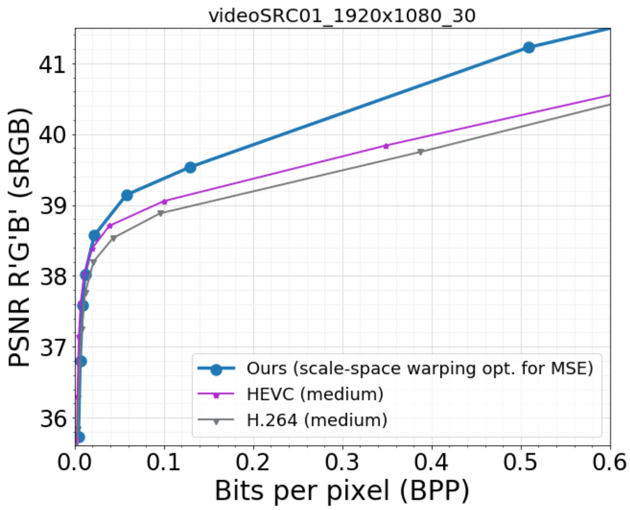


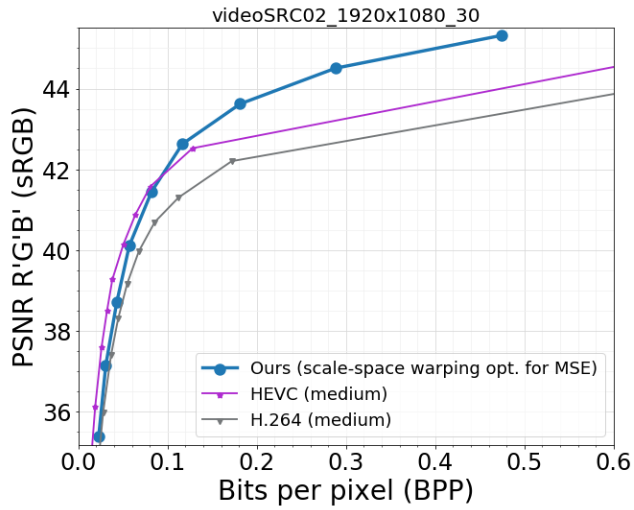

ReadySetGo_1920x1080_120fps_420_8bit_YUV
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
HEVC (very-fast)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

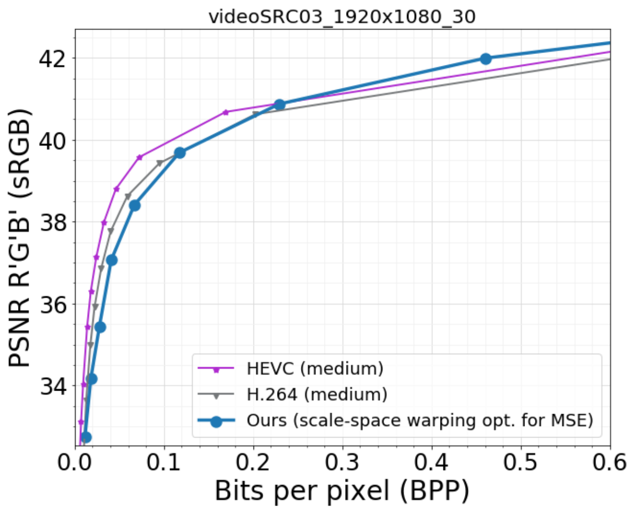

ShakeNDry_1920x1080_120fps_420_8bit_YUV
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)
HEVC (very-fast)

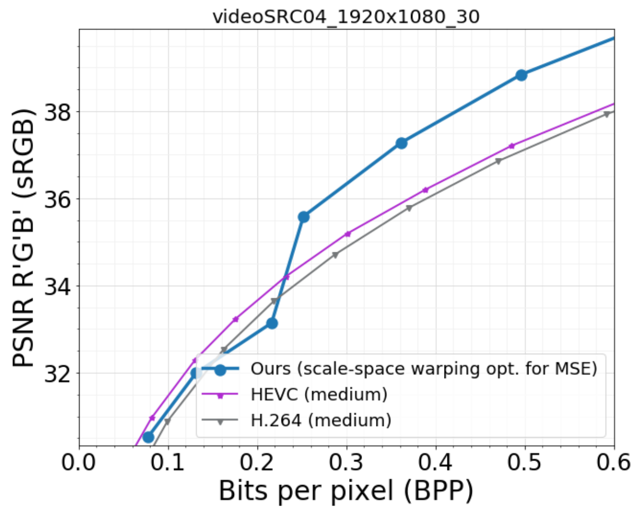

YachtRide_1920x1080_120fps_420_8bit_YUV
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
HEVC (very-fast)
H.264 (medium)

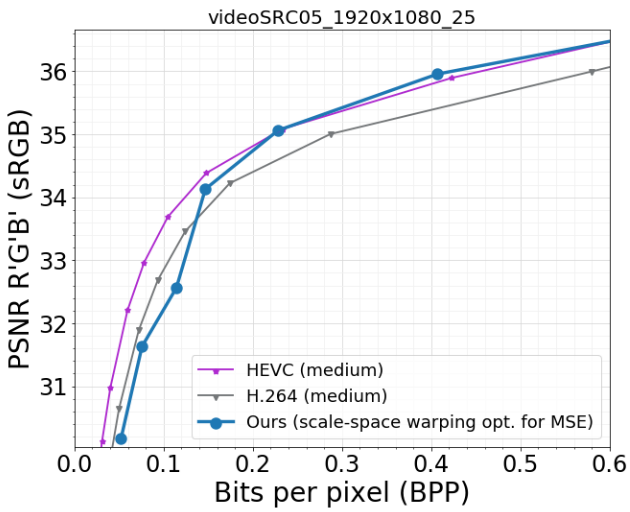
6.2. MCL-JCV

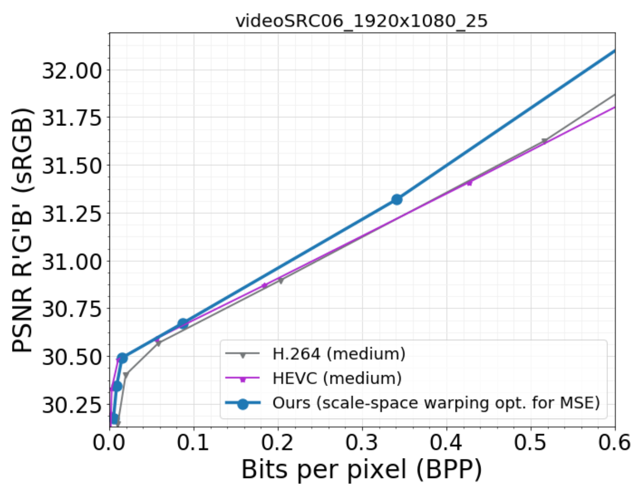

videoSRC01_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

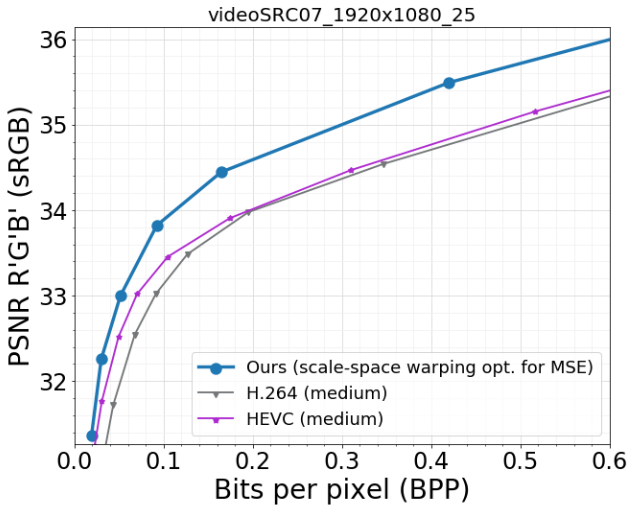

videoSRC02_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

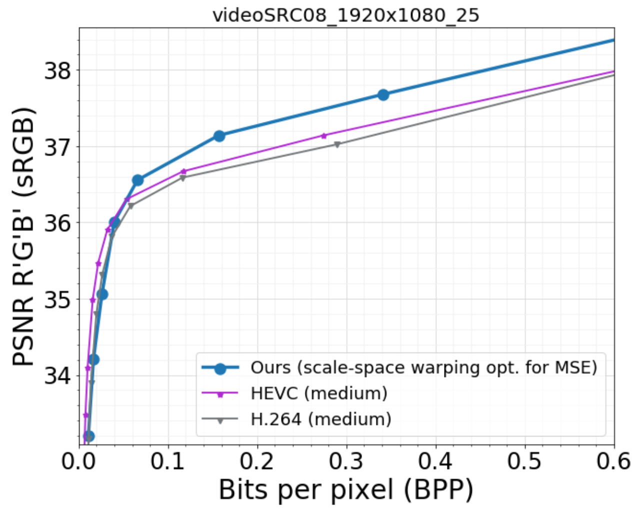

videoSRC03_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

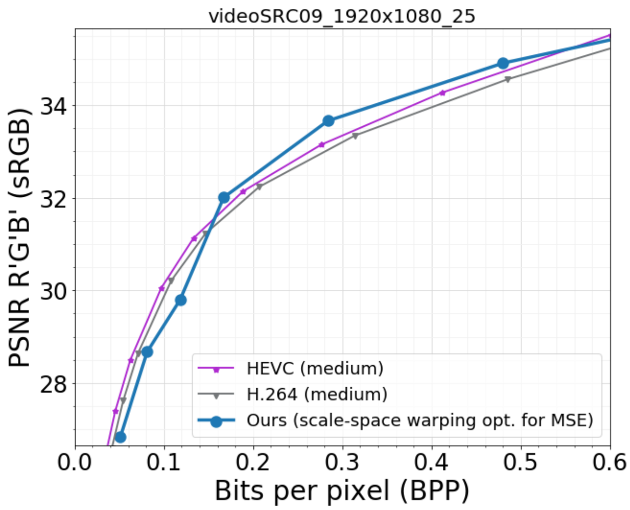

videoSRC04_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

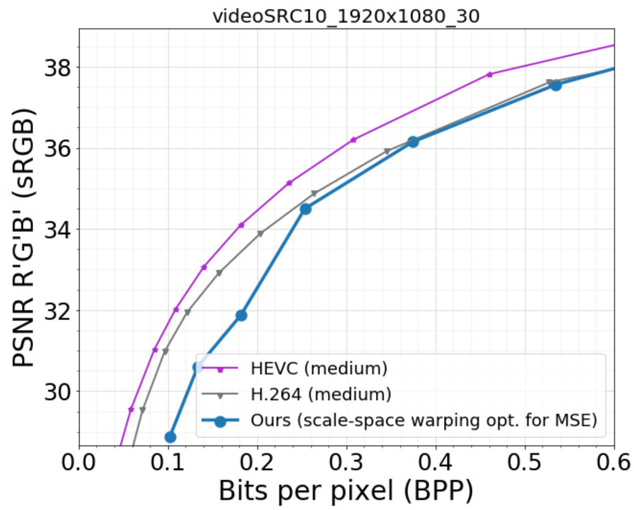

videoSRC05_1920x1080_25
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

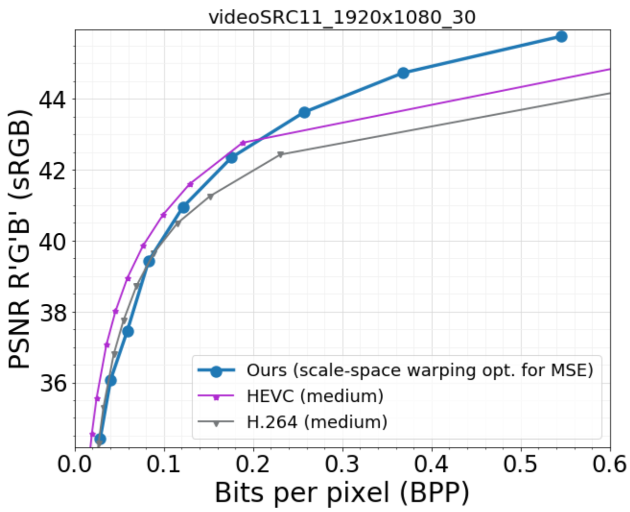

videoSRC06_1920x1080_25
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
H.264 (medium)
HEVC (medium)
Ours (scale-space warping opt. for MSE)

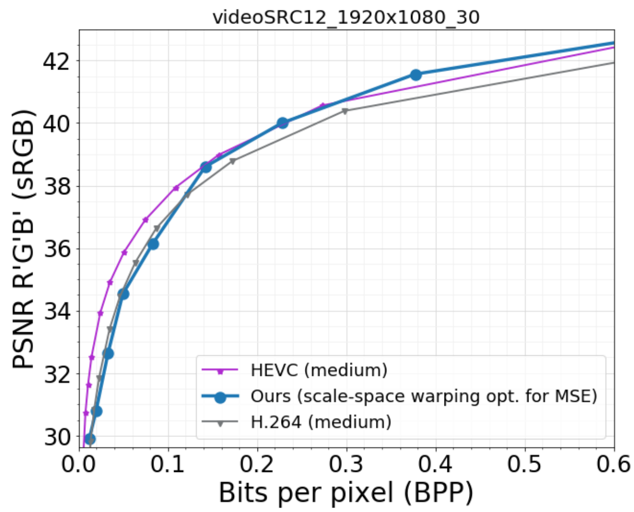

videoSRC07_1920x1080_25
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
H.264 (medium)
HEVC (medium)

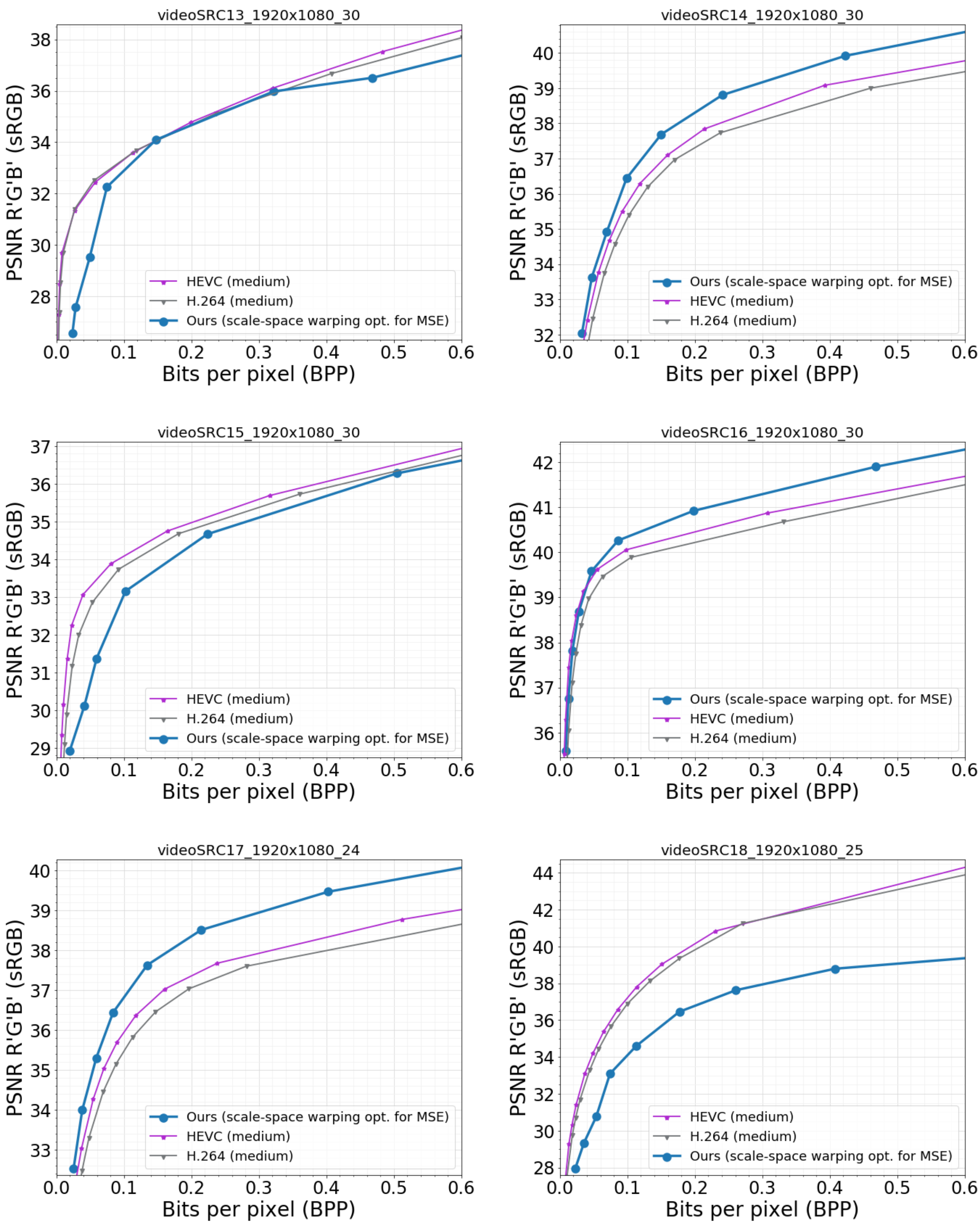

videoSRC08_1920x1080_25
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

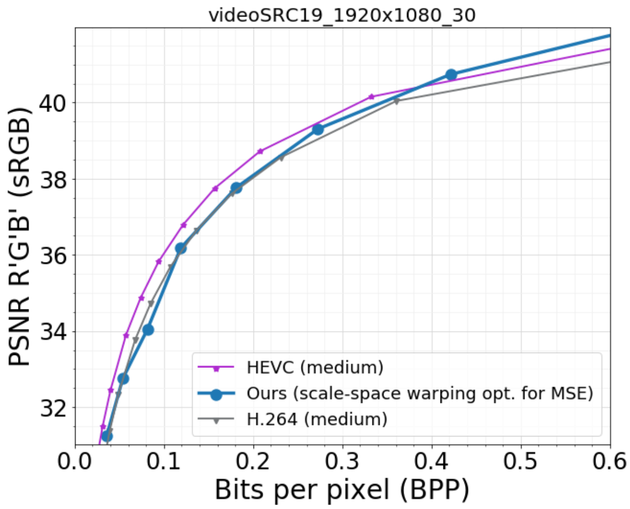

videoSRC09_1920x1080_25
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

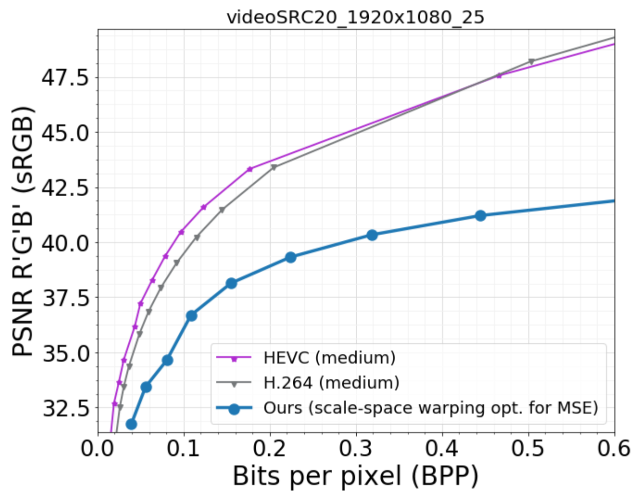

videoSRC10_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

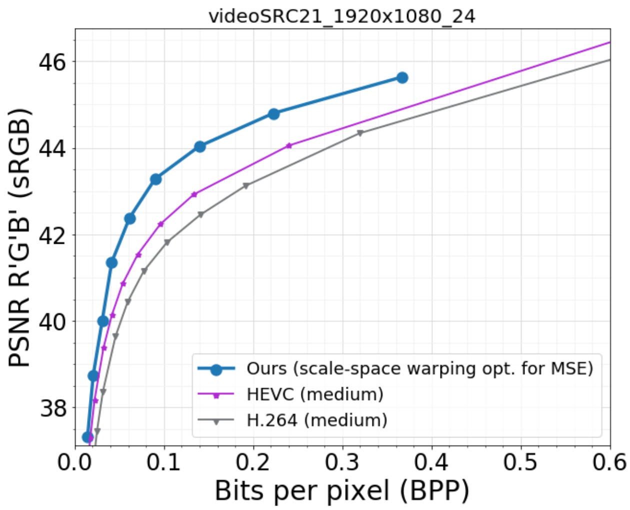

videoSRC11_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

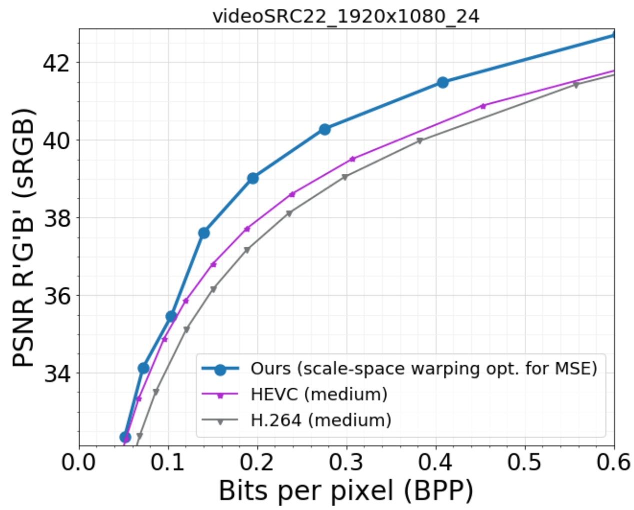

videoSRC12_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
Ours (scale-space warping opt. for MSE)
H.264 (medium)

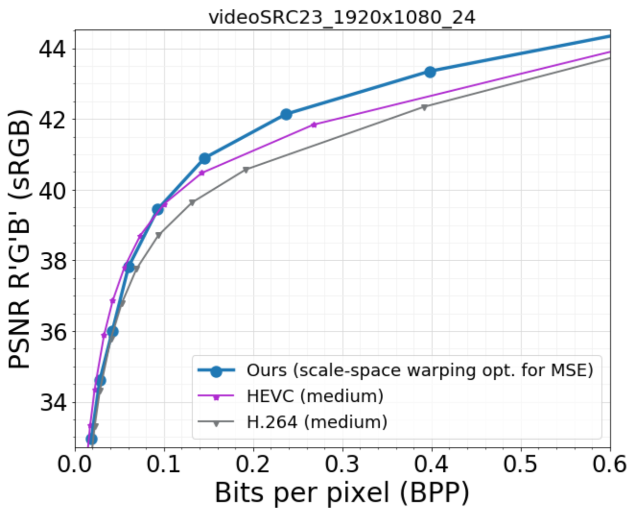

videoSRC13_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)
videoSRC14_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)
videoSRC15_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)
videoSRC16_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)
videoSRC17_1920x1080_24
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)
videoSRC18_1920x1080_25
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

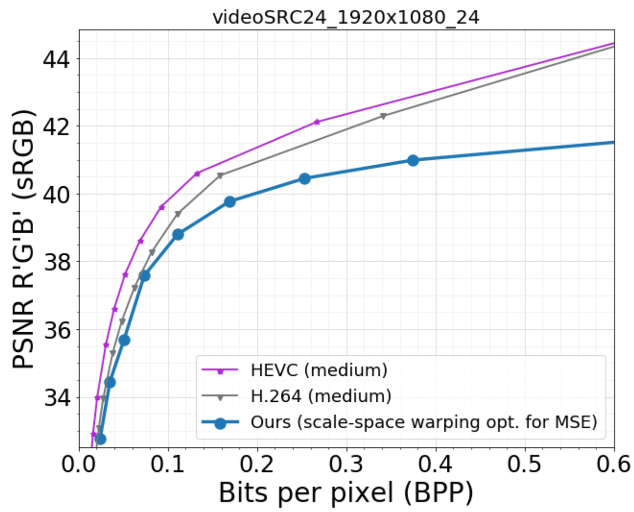

videoSRC19_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
Ours (scale-space warping opt. for MSE)
H.264 (medium)

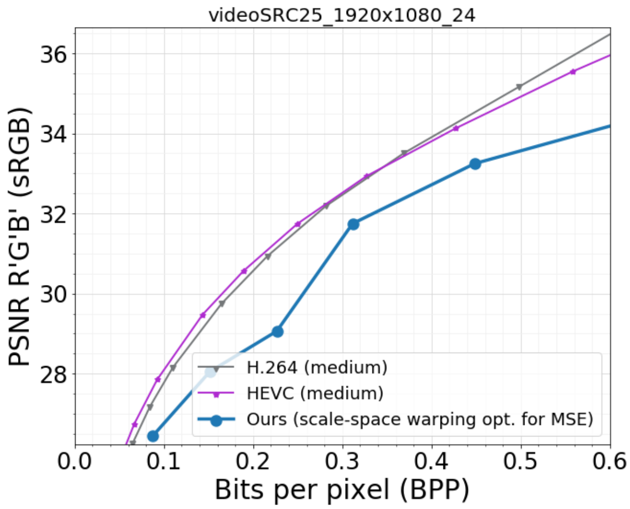

videoSRC20_1920x1080_25
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

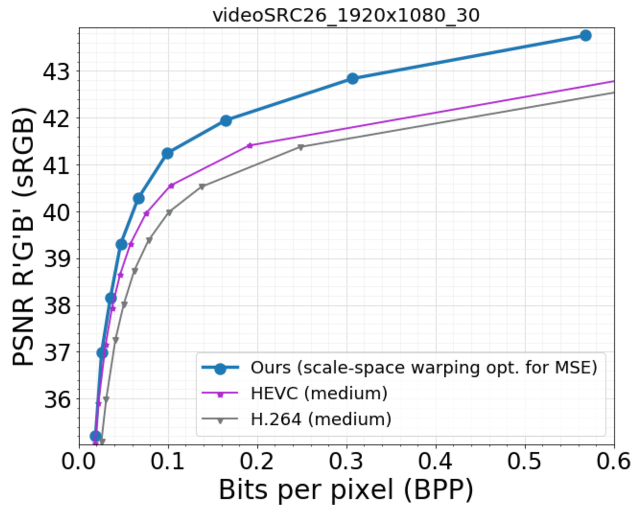

videoSRC21_1920x1080_24
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

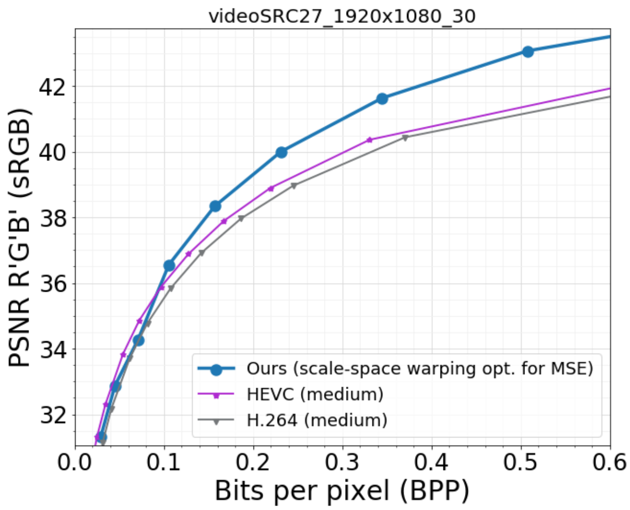

videoSRC22_1920x1080_24
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

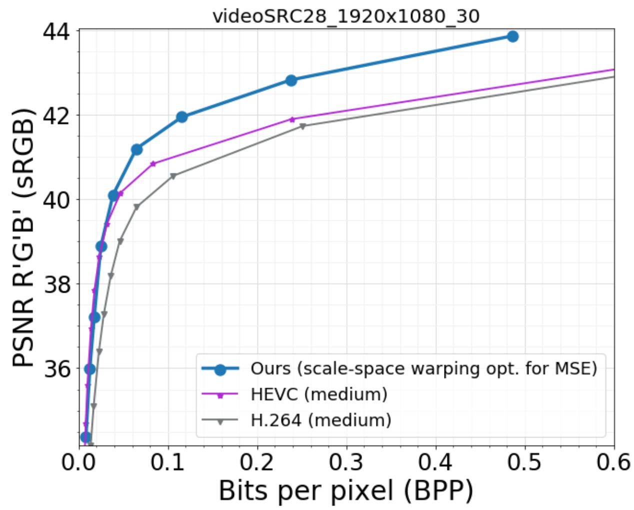

videoSRC23_1920x1080_24
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)

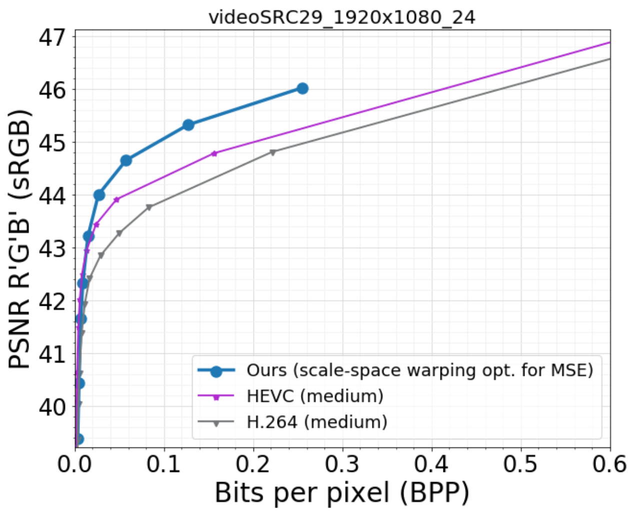

videoSRC24_1920x1080_24
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
H.264 (medium)
Ours (scale-space warping opt. for MSE)

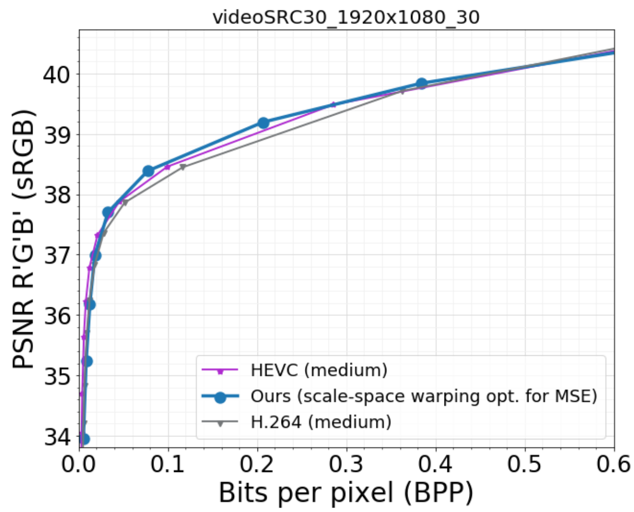

videoSRC25_1920x1080_24
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
H.264 (medium)
HEVC (medium)
Ours (scale-space warping opt. for MSE)


videoSRC26_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)


videoSRC27_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)


videoSRC28_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)


videoSRC29_1920x1080_24
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
Ours (scale-space warping opt. for MSE)
HEVC (medium)
H.264 (medium)


videoSRC30_1920x1080_30
PSNR R'G'B' (sRGB)
Bits per pixel (BPP)
HEVC (medium)
Ours (scale-space warping opt. for MSE)
H.264 (medium)

References

[1] FFmpeg. http://www.ffmpeg.org/. 1

[2] Johannes Ballé, Nick Johnston, and David Minnen. Integer networks for data compression with latent-variable models. 2018. 2

[3] Gisle Bjøntegaard. Calculation of average PSNR differences between RD-curves. Doc. VCEG-M33, ITU-T SG16/Q6 VCEG, Austin, TX, USA, Apr. 2001. 2

[4] Guo Lu, Wanli Ouyang, Dong Xu, Xiaoyun Zhang, Chunlei Cai, and Zhiyong Gao. DVC: An end-to-end deep video compression framework. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11006–11015, 2019. 2