Scale-space flow for end-to-end optimized video compression

Anonymous CVPR submission

Paper ID 9692

Contents

1. Baselines 1
2. Architecture 2
3. Calculating the Bjøntegaard Delta Bit Rate 2
4. Subjective Comparison 4
5. Aggregate Rate-Distortion Graphs 6
6. Per-Video Rate-Distortion Graphs 12

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)

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

HEVC (medium)

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

<table>
<thead>
<tr>
<th>Hyper Encoder</th>
<th>Hyper Decoder (scale)</th>
<th>Hyper Decoder (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv, 5x5/2, 192ch</td>
<td>DeConv, 5x5/2, 192ch</td>
<td>DeConv, 5x5/2, 192ch</td>
</tr>
<tr>
<td>ReLu</td>
<td>QReLu</td>
<td>ReLu</td>
</tr>
<tr>
<td>Conv, 5x5/2, 192ch</td>
<td>DeConv, 5x5/2, 192ch</td>
<td>DeConv, 5x5/2, 192ch</td>
</tr>
<tr>
<td>ReLu</td>
<td>QReLu</td>
<td>ReLu</td>
</tr>
<tr>
<td>Conv, 5x5/2, 192ch</td>
<td>DeConv, 5x5/2, 192ch</td>
<td>DeConv, 5x5/2, 192ch</td>
</tr>
<tr>
<td>QReLu</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

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

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).
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)

![UVG (1080p) graph]

- Ours (scale-space warping opt. for MS-SSIM)
- Ours (scale-space warping opt. for MSE)
- HEVC (medium)
- Habibian 2019 [16]
- H.264 (medium)
- HEVC (very-fast)
- DVC (Lu 2019) [21]
- Wu 2018 [35]
5.3. MCL-JCV (PSNR)

![Graph showing performance comparison between different encoding methods.

- HEVC (medium)
- Ours (scale-space warping opt. for MSE)
- H.264 (medium)
- Ours (scale-space warping opt. for MS-SSIM)
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)

![Graph showing MCL-JCV (1080p) performance with various bit rates and methods]

- **Ours (scale-space warping opt. for MS-SSIM)**
- **Ours (scale-space warping opt. for MSE)**
- **HEVC (medium)**
- **H.264 (medium)**

Bits per pixel (BPP) vs. MS-SSIM RGB (sRGB) for different methods.
5.4.1 Only “Natural” Videos in MCL-JCV (MS-SSIM)

![Graph showing MS-SSIM results for natural videos]

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

![Graph showing MS-SSIM results for animated videos]
6. Per-Video Rate-Distortion Graphs

6.1. UVG

![Per-Video Rate-Distortion Graphs](image)
6.2. MCL-JCV

- videoSRC01_1920x1080_30
- videoSRC02_1920x1080_30
- videoSRC03_1920x1080_30
- videoSRC04_1920x1080_30
- videoSRC05_1920x1080_25
- videoSRC06_1920x1080_25

PSNR R’G’B’ (sRGB) vs. Bits per pixel (BPP)

- Ours (scale-space warping opt. for MSE)
- HEVC (medium)
- H.264 (medium)
videoSRC19_1920x1080_30

videoSRC20_1920x1080_25

videoSRC21_1920x1080_24

videoSRC22_1920x1080_24

videoSRC23_1920x1080_24

videoSRC24_1920x1080_24
References


