

1080

A. Softmax and Entropy

1081

Theorem 1. $H(\{p_0, p_1, p_2\})$ is a monotonic decreasing function of β , where

1082

$$p_i = \frac{e^{\beta y_i}}{\sum_{j=0}^2 e^{\beta y_j}}, \quad i = 0, 1, 2, \quad (8)$$

1083

and $\beta > 0$.

1084

Proof. Let $x = e^\beta$ and $A = x^{y_0} + x^{y_1} + x^{y_2}$. Then,

1085

$$-H(\{p_0, p_1, p_2\}) = -H\left(\left\{\frac{x^{y_0}}{A}, \frac{x^{y_1}}{A}, \frac{x^{y_2}}{A}\right\}\right) \quad (9)$$

1086

$$= \frac{x^{y_0}}{A} \log \frac{x^{y_0}}{A} + \frac{x^{y_1}}{A} \log \frac{x^{y_1}}{A} + \frac{x^{y_2}}{A} \log \frac{x^{y_2}}{A}. \quad (10)$$

1087

The derivative of $-H$ with respect to x is given by

1088

$$-\frac{\partial H}{\partial x} = \sum_{i=0}^2 \left(1 + \log \frac{x^{y_i}}{A}\right) \frac{y_i x^{y_0-1} A - x^{y_i} A'}{A^2} \quad (11)$$

1089

$$= \sum_{i=0}^2 \log \frac{x^{y_i}}{A} \times \frac{y_i x^{y_0-1} A - x^{y_i} A'}{A^2} \quad (12)$$

1090

$$= \sum_{i=0}^2 (\log x^{y_i} - \log A) \frac{y_i x^{y_0-1} A - x^{y_i} A'}{A^2} \quad (13)$$

1091

$$= \sum_{i=0}^2 \log x^{y_i} \times \frac{y_i x^{y_0-1} A - x^{y_i} A'}{A^2} \quad (14)$$

1092

$$= \sum_{i=0}^2 \log x \times \frac{y_i (y_i x^{y_0-1} A - x^{y_i} A')}{A^2} \quad (15)$$

1093

where

1094

$$A' = \frac{\partial A}{\partial x} = y_0 x^{y_0-1} + y_1 x^{y_1-1} + y_2 x^{y_2-1}. \quad (16)$$

1095

Note that

1096

$$(y_0 x^{y_0-1} + y_1 x^{y_1-1} + y_2 x^{y_2-1}) A \\ = (x^{y_0} + x^{y_1} + x^{y_2}) A' \quad (17)$$

1097

and the equalities in (12) and (14) come from (17).

1098

Then, we have

1099

$$-\frac{\partial H}{\partial x} \frac{A^2}{\log x} = \sum_{i=0}^2 y_i^2 x^{y_i-1} A - y_i x^{y_i} A' \quad (18)$$

1100

$$= x^{y_0+y_1-1} (y_0 - y_1)^2 \quad (19)$$

1101

$$+ x^{y_1+y_2-1} (y_1 - y_2)^2 \quad (20)$$

1102

$$+ x^{y_2+y_0-1} (y_2 - y_0)^2 \quad (21)$$

1103

$$\geq 0. \quad (22)$$

1104

Thus, if $x > 1$, then $\frac{\partial H}{\partial x} \leq 0$ and H is a strictly monotonic decreasing function of x unless $y_0 = y_1 = y_2$. Moreover, $x = e^\beta$ is a strictly monotonic increasing function of β , and $x > 1$ if $\beta > 0$. Therefore, H is a strictly monotonic decreasing function of β , provided that $\beta > 0$. \square

B. Implementation and Training Details

1134

First, we describe the software for traditional codecs and the libraries for learning-based algorithms. Second, we present the implementation details of the proposed context models CRR and CDR. Then, we explain how to train the proposed CTC algorithm.

1135

B.1. Software and Libraries

1136

We adopt the traditional codecs JPEG2000 [2], BPG444 [10], VTM 12.0 [11] for comparison.

1137

JPEG2000: We use the open software in [2]. We execute the following commands for encoding and decoding. We transform RGB-formatted images, such as png files, into raw files.

1138

```
{buildpath}/opj_compress -i {inputfile} -o {bin} -r {15:150} -F {width},{height},3,8,u@1x1:1x1:1x1
```

1139

```
{buildpath}/opj_decompress -i {bin} -o {outputfile}
```

1140

BPG444: We use the software in [10] and enter the following commands.

1141

```
{buildpath}/bpgenc {inputfile} -o {bin} -q {26:52} -f 444 -e x265
```

1142

```
{buildpath}/bpgdec -o {outputfile} {bin}
```

1143

VTM 12.0: We execute the reference software package in https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM-/tree/VTM-12.0 with the following commands.

1144

```
{buildpath}/EncoderApp -i {inputfile} -c {cfgpath}/encoder_intra_vtm.cfg -o /dev/null -b {bin} -wdt {width} -hgt {height} -fr 1 -f 1
```

1145

```
-q 24, 26, 30, 31:43 --InputChromaFormat=444 --InputBitDepth=8 --ConformanceWindowMode=1
```

1146

```
--InputColourSpaceConvert=RGBtoGBR --SNRInternalColourSpace=1 --OutputInternalColourSpace=0
```

1147

```
{buildpath}/DecoderApp -b {bin} -o {outputfile} -d 8 --OutputColourSpaceConvert=GBRtoRGB
```

1148

We use Pytorch [36] and CompressAI [8] libraries to implement the proposed CTC algorithm. Also, we employ the source codes and pretrained parameters in CompressAI for the Minnen *et al.*'s algorithm [33]. For

1188 the other learning-based codecs, we use the results provided
 1189 in the original papers.
 1190

B.2. Implementation of CRR and CDR

1192 The main network of the proposed CTC algorithm is in
 1193 Figure 2, and the detailed structures of the CRR and CDR
 1194 modules are in Figure 4. The context modules are incor-
 1195 porated into the main network as follows. There are three
 1196 CRR models for different intervals of trit-plane levels l . We
 1197 denote them as CRR_L , CRR_{L-1} , and $\text{CRR}_{\leq L-2}$, where the
 1198 subscripts indicate the ranges of trit-plane levels in which
 1199 the corresponding models are used. In other words,
 1200

$$\tilde{\mathbf{P}}_l = \begin{cases} \text{CRR}_L(\hat{\mathbf{Y}}_{l-1}, \mathbf{M}, \Sigma, \mathbf{E}_l, \mathbf{P}_l) & \text{if } l = L, \\ \text{CRR}_{L-1}(\hat{\mathbf{Y}}_{l-1}, \mathbf{M}, \Sigma, \mathbf{E}_l, \mathbf{P}_l) & \text{if } l = L - 1, \\ \text{CRR}_{\leq L-2}(\hat{\mathbf{Y}}_{l-1}, \mathbf{M}, \Sigma, \mathbf{E}_l, \mathbf{P}_l) & \text{if } l \leq L - 2. \end{cases} \quad (23)$$

1201 Similarly, we implement three CDR models CDR_{L-1} ,
 1202 CDR_{L-2} , and $\text{CDR}_{\leq L-3}$ to obtain
 1203

$$\tilde{\mathbf{Y}}_l = \begin{cases} \hat{\mathbf{Y}}_l & \text{if } L - 1 < l \leq L, \\ \text{CDR}_{L-1}(\hat{\mathbf{Y}}_l, \mathbf{M}, \Sigma) & \text{if } L - 2 < l \leq L - 1, \\ \text{CDR}_{L-2}(\hat{\mathbf{Y}}_l, \mathbf{M}, \Sigma) & \text{if } L - 3 < l \leq L - 2, \\ \text{CDR}_{\leq L-3}(\hat{\mathbf{Y}}_l, \mathbf{M}, \Sigma) & \text{if } l \leq L - 3. \end{cases} \quad (24)$$

1204 Whereas CRR estimates the probability tensor $\tilde{\mathbf{P}}_l$ for
 1205 each trit-plane \mathbf{T}_l ($l = 1, \dots, L$), CDR performs the pre-
 1206 diction of $\tilde{\mathbf{Y}}_l$ for any $l \leq L - 1$. Therefore, l is an integer
 1207 in (23) but a real number in (24). The trit-plane levels
 1208 $l \leq L - 2$ are supported by a single CRR model, and the
 1209 levels $l \leq L - 3$ are by a single CDR model. These choices
 1210 are made to strike a balance between the number of par-
 1211 ameters and the RD performance. Also, CDR is not used at the
 1212 top level $L - 1 < l \leq L$ because the refinement of a latent
 1213 tensor $\hat{\mathbf{Y}}_l$ is not necessary at such a fine level.
 1214

B.3. Training of CTC

1215 We train the main network for 300 epochs using the rate-
 1216 distortion loss $\mathcal{L} = \mathcal{D} + \lambda \mathcal{R}$ with $\lambda = 5$.
 1217

1218 In the trit-plane slicing module in Figure 2, a latent ten-
 1219 sor is sliced into L trit-planes. Note that the maximum
 1220 trit-plane level L depends on the latent tensor, as described
 1221 in [22]. However, $L = 7$ for most images. The selection
 1222 of CRR and CDR models in (23) and (24) is dependent on
 1223 L . Therefore, for stable training of these models, as well as
 1224 the decoder retraining, we fix $L = 7$ and use the training
 1225 images with $L = 7$ only.
 1226

1227 We use the cross-entropy loss in (4) to train the three
 1228 CRR models. The CRR process is performed for every trit,
 1229 except when the original probabilities are $(p_0, p_1, p_2) =$
 1230

1231 $(0, 1, 0)$. In such a case, the trit requires no bit, and there
 1232 is no reason to update its probabilities.
 1233

1234 The CDR loss in (6) can be rewritten as
 1235

$$\ell_{\text{CDR}}(l) = \|\mathbf{Y} - \tilde{\mathbf{Y}}_l\|_F \quad (25)$$

1236 where l denotes a trit-plane level. The first CDR model
 1237 CDR_{L-1} in (24) supports a partially reconstructed level
 1238 $l \in (L - 2, L - 1]$. For its training, we use the sum of
 1239 losses, given by
 1240

$$\ell_{\text{CDR}}(L - 1) + \ell_{\text{CDR}}(\alpha) + \ell_{\text{CDR}}(L - 2) \quad (26)$$

1241 where $\alpha \sim \mathcal{U}(0, 1)$ is a uniform random variable. The
 1242 losses for the other two models CDR_{L-2} and $\text{CDR}_{\leq L-3}$
 1243 are similarly defined.
 1244

1245 Then, the decoder is retrained to minimize the loss ℓ_{DEC}
 1246 in (7). Note that the original decoder is optimized for the
 1247 case when all trit-planes are received (*i.e.* the highest level
 1248 $l = L$). Thus, the decoder is retrained to consider lower
 1249 levels as well. However, due to the retraining, the per-
 1250 formances at high levels can be degraded. To alleviate the
 1251 degradation, we set large weighting parameters at high lev-
 1252 els, compared to low levels. Specifically, we define the loss
 1253 as
 1254

$$\ell_{\text{DEC}} = 100 \times \sum_{l=L-1}^L \|g_s(\tilde{\mathbf{Y}}_l) - \mathbf{X}\|_F \quad (27)$$

$$+ \sum_{l=L-4}^{L-2} \|g_s(\tilde{\mathbf{Y}}_l) - \mathbf{X}\|_F. \quad (28)$$

1255 Note that we consider five levels from $L - 4$ to L , and set
 1256 bigger weights at the two highest levels L and $L - 1$.
 1257

1258 The training epochs for the context models and the de-
 1259 coder retraining are summarized in Table 4. These training
 1260 schedules are determined by observing the convergence of
 1261 the validation performance.
 1262

C. More Experiments

C.1. RD curves

1263 Figures 12 and 13 compare the RD curves on the CLIC
 1264 validation dataset and the JPEG-AI testset, respectively. All
 1265 learning-based algorithms, including the proposed CTC, are
 1266 optimized to minimize the MS-SSIM loss in Figure 12(b)
 1267 and Figure 13(b).
 1268

1269 Figure 14 compares the proposed CTC with the trit-plane
 1270 coding without RD priorities. More specifically, in ‘Without
 1271 RD priorities,’ the trits in each trit-plane are transmitted
 1272 in the 3D raster scan order, instead of the decreasing order of
 1273 their RD priorities [27]. This alternative method performs
 1274 badly compared to CTC. However, we see that its per-
 1275 formance is also improved by employing the two context mod-
 1276 ules, CRR and CDR, and the decoder retraining scheme.
 1277

1296 Table 4. The numbers of epochs for the context model training and the decoder retraining (g_s). 1350
1297 1351
1298 1352
1299 1353
1300 1354
1301 1355
1302 **C.2. Reconstructed images of CTC** 1356
1303 1357
1304 Figures 15~20 show various images reconstructed by 1358
1305 the proposed CTC algorithm at levels $l = L, L - 2, L - 3,$ 1359
1306 and $L - 4$. The images with resolutions larger than 512×768 1360
1307 are center-cropped. 1361
1308 1362
1309 1363
1310 1364
1311 1365
1312 1366
1313 1367
1314 1368
1315 1369
1316 1370
1317 1371
1318 1372
1319 1373
1320 1374
1321 1375
1322 1376
1323 1377
1324 1378
1325 1379
1326 1380
1327 1381
1328 1382
1329 1383
1330 1384
1331 1385
1332 1386
1333 1387
1334 1388
1335 1389
1336 1390
1337 1391
1338 1392
1339 1393
1340 1394
1341 1395
1342 1396
1343 1397
1344 1398
1345 1399
1346 1400
1347 1401
1348 1402
1349 1403

| CRR _L | CRR _{L-1} | CRR _{<L-2} | CDR _{L-1} | CDR _{L-2} | CDR _{<L-3} | g_s |
|------------------|--------------------|------------------------|--------------------|--------------------|------------------------|-------|
| 300 | 300 | 10 | 30 | 30 | 30 | 100 |

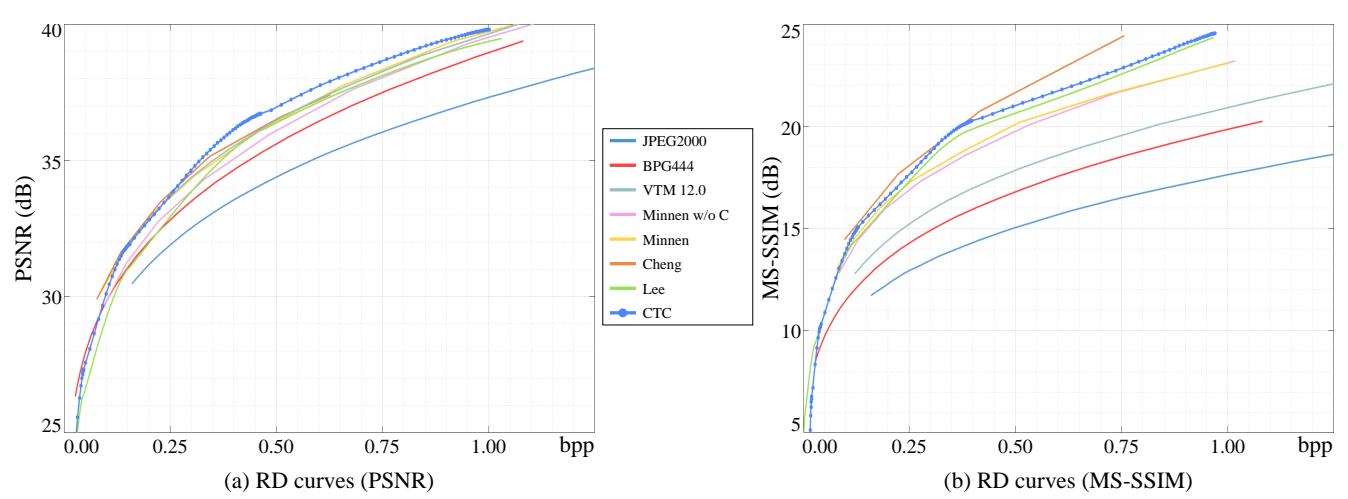


Figure 12. RD curve comparison on the CLIC validation dataset.

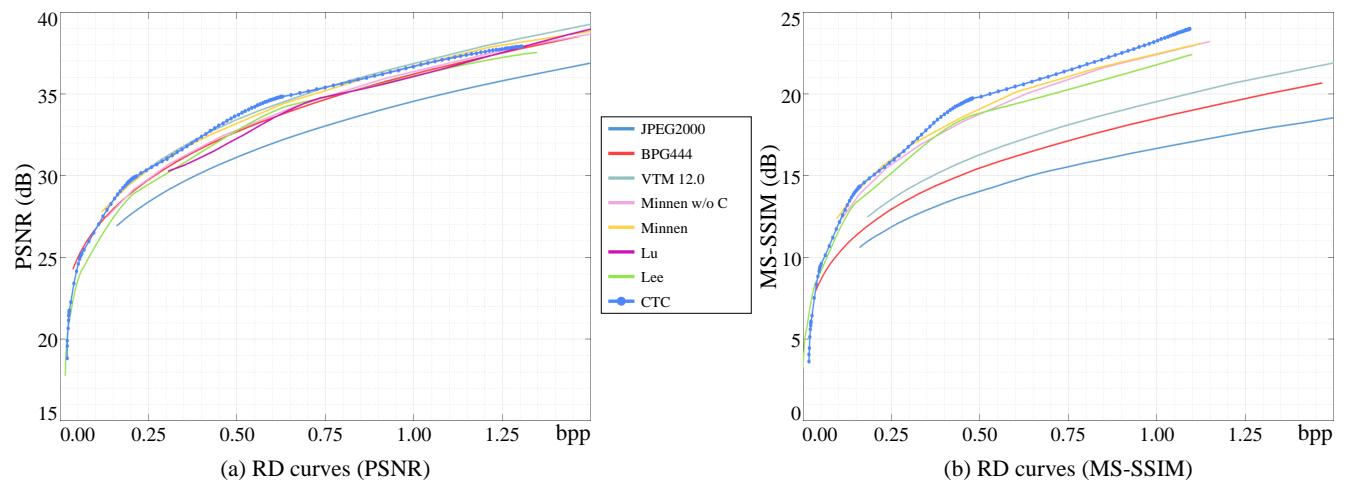


Figure 13. RD curve comparison on the JPEG-AI testset.

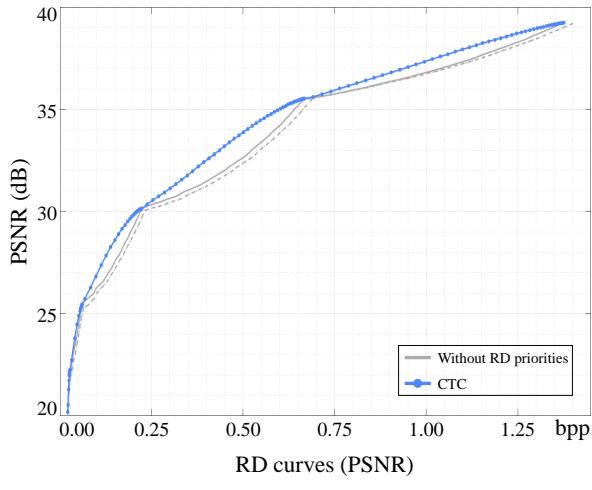


Figure 14. RD curve comparison of CTC and the alternative trit-plane coding method without RD priorities on the Kodak lossless dataset. The dashed curve means that the context modules, CRR and CDR, and the decoder retraining are not employed.

1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565



Figure 15. Reconstructed images of “kodim01.png” and “kodim02.png.” The bitrates (bpp) and PSNRs (dB) are reported below each image.



1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781



Figure 17. Reconstructed images of “ales-krivec15949.png” and “andrew-ruiz-376.png.” The bitrates (bpp) and PSNRs (dB) are reported below each image.

1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889

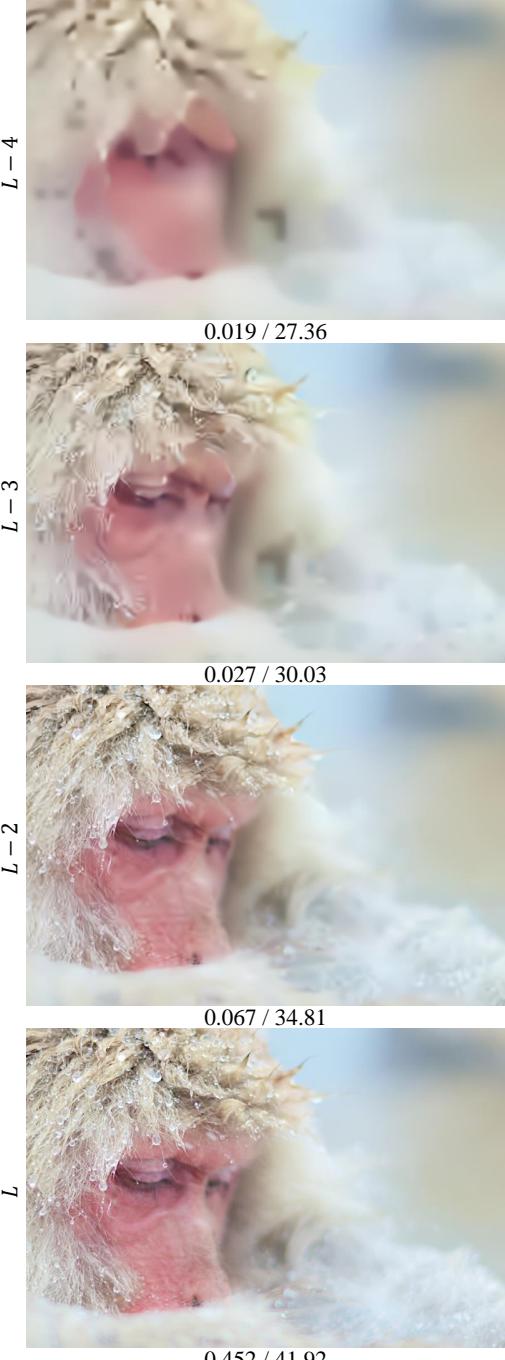


Figure 18. Reconstructed images of “nomao-saeki-33553.png” and “philipp-reiner-207.png.” The bitrates (bpp) and PSNRs (dB) are reported below each image.

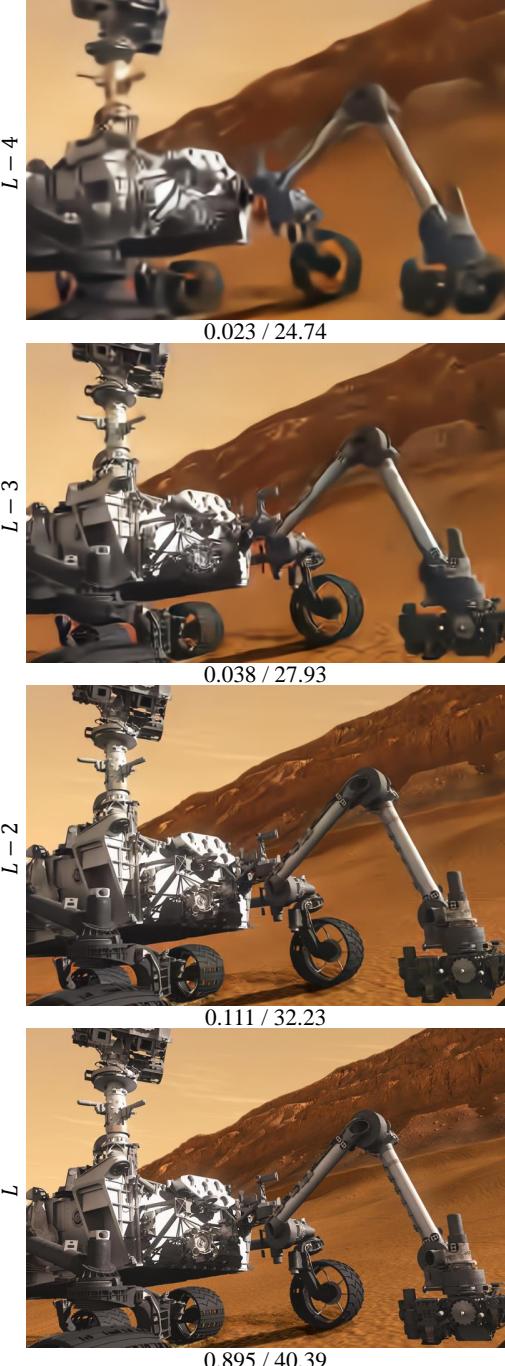
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997



Figure 19. Reconstructed images of “000505_TE_1336x872.png” and “000505_TE_1336x872.png.” The bitrates (bpp) and PSNRs (dB) are reported below each image.

1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051

2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105



L - 4
0.023 / 24.74

L - 3
0.038 / 27.93

L - 2
0.111 / 32.23

L
0.895 / 40.39



L - 4
0.024 / 22.66

L - 3
0.049 / 26.47

L - 2
0.145 / 31.59

L
0.824 / 41.33

2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159

Figure 20. Reconstructed images of “00010_TE_2000x1128.png” and “00015_TE_3680x2456.png.” The bitrates (bpp) and PSNRs (dB) are reported below each image.