JEG-Hybrid

Framework for reproducible objective video quality research with case study on PSNR implementations

AHMED ALDAHDOOH
ENRICO MASALA
GLENN VAN WALLENDAEL
MARCUS BARKOWSKY
Reproducibility

- algorithms are complex,
- depend on specific implementations in software packages
- their parameters need to be trained on a particular dataset

- Textual descriptions:
  - lack the required detail
  - even for the simple Peak Signal to Noise Ratio (PSNR)
Reproducibility: a sample example

- Search three terms in IEEEXplore: video, quality, and prediction
- 59 hits
- 16: no new algorithm → discarded
Reproducibility: a sample example

- Green: 9.31% out of the 43 papers: comes with source code allowing reproducibility
- Yellow: 39.53% of the works seem to provide reasonably detailed information
- Red: 51.16% rely on some sort of learning technique (no dataset, no code)
Reproducibility: PSNR

- peak value
  - ITU-R BT.601: 235
  - 8 bit: 255

- temporal alignment
  - overestimates the quality when ignoring
    - Stalling
    - Skipping
    - Reduced frame rate

- color alignment
  - brightness, contrast, and color changes may even improve quality

- temporal pooling
  - averaging the Mean Squared Error (MSE)
  - averaging the PSNR values per frame
  - squared mean of the PSNR values: emphasizing degradations

\[
20 \log_{10} \frac{235}{255} \approx -0.71
\]
Temporal pooling

\[
MSE_f = \sum_{i=1}^{X} \sum_{j=1}^{Y} (\hat{p}_{ij} - p_{ij})^2
\]
\[
PSNR_f = 10 \log_{10} \frac{\text{peak}^2}{MSE_f}.
\]

- **PSNR}_A (arithmetic mean): MSE}_f is averaged over all frames
- **PSNR}_G (geometric mean): the mean of the PSNR}_f

\[
PSNR_A = 10 \log_{10} \frac{255^2}{\frac{1}{N} \sum_{k=1}^{N} PSNR_{f_k}}
\]
\[
PSNR_G = \frac{1}{N} \sum_{k=1}^{N} PSNR_{f_k}.
\]
Impact of temporal pooling

Difference in PSNR up to 4dB depending on temporal pooling.

Mostly below 0.5dB
Impact of temporal pooling

- Only considering rate control:
Variance as an extra indicator

Higher $\sigma^2_{PSNR}$ yields to higher difference

if $\sigma^2_{PSNR}$ it is lower than 2
$PSNR$ difference < 0.5 dB

if $\sigma^2_{PSNR}$ it is higher than 4
$PSNR$ difference can be 1.5 dB
Variance as an extra indicator

Higher $\sigma^2_{\text{PSNR}}$ yields to higher difference

if $\sigma^2_{\text{PSNR}}$ it is lower than 2
PSNR difference < 0.5 dB
Variance as an extra indicator

Higher $\sigma^2_{\text{PSNR}}$ yields to higher difference

if $\sigma^2_{\text{PSNR}}$ it is lower than 2
PSNR difference < 0.5 dB

if $\sigma^2_{\text{PSNR}}$ it is higher than 4
PSNR difference can be 1.5 dB
JEG-Hybrid

Subset selection for the design of subjective experiments

AHMED ALDAHDOOH
ENRICO MASALA
GLENN VAN WALLENDAEL
MARCUS BARKOWSKY
Subset selection

- You do not want to select
  - different quality levels, i.e. different QPs
  - different bitrate budgets, nor
  - selecting different content types

- choose the HRCs that cover a wide range of the targets
Subset selection

- **Quality/Bitrate-driven HRCs Subset**
  - representing all ranges of PSNR and bitrates

- **Content-driven HRCs Subset**
  - algorithm for selecting the HRCs that behave differently with the contents
Goal-driven Large-scale Database Subset Generation

- predict the behavior of a full-reference quality measure (VQM)

<table>
<thead>
<tr>
<th>Trained on</th>
<th>Content-based</th>
<th>Quality/bitrate-based</th>
<th>Random 1</th>
<th>Random 2</th>
<th>Random 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content-based</td>
<td>0.99</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Quality/bitrate</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 1</td>
<td>0.98</td>
<td>0.96</td>
<td>0.99</td>
<td>0.97</td>
<td>0.90</td>
</tr>
<tr>
<td>Random 2</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Random 3</td>
<td>0.63</td>
<td>0.60</td>
<td>0.69</td>
<td></td>
<td>0.92</td>
</tr>
</tbody>
</table>