Deep Neural Network Based Frame Reconstruction For Optimized Video Coding - An AV2 Approach

Dandan Ding
Hangzhou Normal University
AV1 is the most advanced standardized codec available today.

Research and development of tools towards a potential successor to AV1, so called AV2, have started.

A viable successor for further BDRATE reduction over AV1.
Our Goal

We completely focus on the optimization of reconstruction frames through using the Deep Neural Network (DNN).
Two problems are concerned

Two aspects are explored, including:

Q1 How to design a CNN-based in-loop filter for AV1?

Q2 How to incorporate the CNN-based filters into AV1 encoder?
How to design a CNN-based in-loop filter for AV1?

- The problem has similarities with the SR problem.

Loss function: \[ L(\Theta) = \frac{1}{n} \sum_{i=1}^{n} ||F(Y_i; \Theta) - X_i||^2 \]

process the in-loop filter in the same way.

Raw reconstructed frame \hspace{1cm} CNN network \hspace{1cm} Enhanced reconstructed frame

Classical CNNs

Test conditions:
- HM 16.9
- 18 images
- QP=37
- Intra coding
- The anchor in-loop filters are turned off


Performance of the CNN-based in-loop filtering

The PSNR gain is as large as 0.8dB.
But using large amount of parameters is expensive!

Test conditions
- AV1 platform (Sept.)
- 18 images
- QP=53
- Only intra coding

To obtain a slim version
- Reduces the number of channels
- Reduce the kernel size
- Select a balanced number of layers

0.25dB can be achieved with 20k parameters.
How to incorporate the CNN-based filters into video encoders?

- Previous work focuses on designing various CNN structures.
- These CNNs are directly incorporated into encoders for in-loop filtering.
How to incorporate the CNN-based filters into video encoders?

- The filtered frames will be referenced in the subsequent coding.
- Then can more gains be expected from inter coding?

The over-filtering problem in AV1 inter (left), HEVC LDP (middle), and HEVC RA (right)
The test condition is inconsistent with the training condition.

- We conduct end-to-end training and obtain a model, without considering the intertwined correlations across frames.
- But there exists complex reference relationships in practical coding.

How to avoid the over-filtering problem?

Such a “Direct” training obtains a locally optimal model.

- A direct replacement using the “direct” model will trigger over-filtering problem.
- We cannot obtain a global optimum model because it is impossible to simulate the correlations across frame in coding.
Some remedies to redress the over-filtering problem

**Solution 1**

**01 Rate-Distortion method**

02 Skipping method

Only apply CNN to selective regions or frames
Results on AV1

- Only frame 2, 6, 10 and 14 are filtered by CNN.
- Around 0.22dB gain is retained.

<table>
<thead>
<tr>
<th>Frame No.</th>
<th>AV1 Anchor</th>
<th>Enhance every frame</th>
<th>Proposed skipping strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>31.40</td>
<td>31.78</td>
<td>31.78</td>
</tr>
<tr>
<td>1</td>
<td>29.80</td>
<td>29.68</td>
<td>29.96</td>
</tr>
<tr>
<td>2</td>
<td>29.75</td>
<td>29.67</td>
<td>29.89</td>
</tr>
<tr>
<td>3</td>
<td>29.33</td>
<td>29.01</td>
<td>29.45</td>
</tr>
<tr>
<td>4</td>
<td>29.70</td>
<td>29.71</td>
<td>29.86</td>
</tr>
<tr>
<td>5</td>
<td>29.14</td>
<td>28.95</td>
<td>29.37</td>
</tr>
<tr>
<td>6</td>
<td>29.34</td>
<td>29.18</td>
<td>29.61</td>
</tr>
<tr>
<td>7</td>
<td>29.27</td>
<td>29.08</td>
<td>29.57</td>
</tr>
<tr>
<td>8</td>
<td>29.95</td>
<td>30.05</td>
<td>30.13</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td><strong>29.74</strong></td>
<td><strong>29.68</strong></td>
<td><strong>29.96</strong></td>
</tr>
</tbody>
</table>

Dandan Ding, Guangyao Chen, Debargha Mukherjee, Urvang Joshi, and Yue Chen, A CNN-based in-loop filtering approach for AV1 video codec, PCS, 2019.

Guangyao Chen, Dandan Ding, Debargha Mukherjee, Urvang Joshi, and Yue Chen, AV1 in-loop filtering using a wide-activation structured residual network, IEEE ICIP, 2019.
Visual quality

(a) Anchor  (b) Apply CNN to every frame

(c) CTU-RDO  (d) Skipping method
Solution 2

Train a global model

- Fundamentally solve the over-filtering problem.
- We propose a progressive training method.
  - Through transfer learning, the reconstructed frames that have been filtered by the CNN models are progressively involved back to fine-tune the CNN models themselves.

Raw reconstructed frame → Training set → CNN network → Enhanced reconstructed frame
Visual quality

Original frame

CTU-RDO

Proposed global model
Results of our global model

- The global model can further improve the performance of RDO.
- A direct application of the global model to each frame will achieve a comparable gain to that of RDO.

Different solutions for over-filtering problem (PSNR)

<table>
<thead>
<tr>
<th>CTU-RDO using Direct model</th>
<th>CTU-RDO using the global model</th>
<th>Directly applying the global model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitrate</td>
<td>PSNR</td>
<td>Bitrate</td>
</tr>
<tr>
<td>951.36</td>
<td>32.74</td>
<td>952.21</td>
</tr>
</tbody>
</table>

Test conditions

- HEVC: HM16.9
- QP=37
- 50 inter frames
- RA configuration
Multi-frame video enhancement

• Above studies are all on basis of single frame.
• Videos introduce an additional time dimension.
• How to utilize the information from temporal domain?

• There is frame-level quality fluctuation in compressed videos.
• A pair of high-quality frames can be utilized to enhance the low-quality frames in between.

Results on AV1


Performance of multi-frame method on AV1 (PSNR)

<table>
<thead>
<tr>
<th></th>
<th>AV1 anchor</th>
<th>Multi frame</th>
<th>Single frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter</td>
<td>35.49</td>
<td><strong>35.84</strong></td>
<td>35.71</td>
</tr>
<tr>
<td>Intra</td>
<td>30.72</td>
<td><strong>31.57</strong></td>
<td>30.86</td>
</tr>
</tbody>
</table>

Test conditions

- QP=53
- Only 36 low-quality frames
- Flownet2.0 is employed for motion estimation
Conclusion

• Two problems are concerned when embedding the CNN-based tools into video encoders.
  • The CNN structure
  • The incorporation approaches

• Currently, we employ a single CNN model to deal with all videos.

• It is possible to develop different small CNNs for different video characteristics.
Thank You

DandanDing@hznu.edu.cn
https://github.com/IVC-Projects