Video Enhancement Network Based on Max-pooling and Hierarchical Feature Fusion
Background

Video Enhancement

Figure 1: In-loop filtering and post-processing in Codec
## Background

### Analysis (Learning-base Method)

<table>
<thead>
<tr>
<th>Name</th>
<th>In-loop filtering</th>
<th>Post-processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>It is stable and can be integrated into the encoder, but it will increase the complexity of the encoder and the encoding time.</td>
<td>It does not increase the complexity of the encoder, but does not make full use of the image feature information</td>
</tr>
</tbody>
</table>
Proposed Method

Overview of the Network
Proposed Method

Feature Extraction

Diagram:
- Conv
- HxWxC
- Up-sample
- Concatenate
- MP: Max-pooling
- Up-sample: nearest
Proposed Method

Hierarchical Feature Fusion

In order to make full use of different scale features, we design a hierarchical feature fusion module to fuse multi-scale features.(Stage 1 and stage 2 in the figure)
Proposed Method

Training

• The dataset in this paper comes from 16 HEVC test sequences. All the data are compressed by HEVC reference software (HM) with the configuration QP = 42, All Intra mode. We only use luminance component Y for training and testing.

• In training stage, we decomposed each compressed frame into 64x64 patches. In the testing stage, we take the whole frame as input. The purpose of this design of dataset is to obtain a large number of samples with different content features. At the same time, accelerating the convergence of the model in the training stage.

• We train the model with Adam optimizer, learning rate 1e-4, batch size 64, and mean square error as loss function. The deep learning framework is Pytorch 1.1.0.
Result

Compared with HEVC(PSNR)

<table>
<thead>
<tr>
<th></th>
<th>BasketballDrill ΔPSNR/dB</th>
<th>BasketballPass ΔPSNR/dB</th>
<th>Johnny ΔPSNR/dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>VRCNN[20]</td>
<td>0.4104</td>
<td>0.2911</td>
<td>0.4734</td>
</tr>
<tr>
<td>DSCNN[13]</td>
<td>0.4033</td>
<td>0.2621</td>
<td>0.4988</td>
</tr>
<tr>
<td>DCAD[12]</td>
<td>0.4486</td>
<td>0.3208</td>
<td>0.5573</td>
</tr>
<tr>
<td>Ours</td>
<td>0.4861</td>
<td>0.3415</td>
<td>0.5613</td>
</tr>
</tbody>
</table>

Results of test 2. ΔPSNR denotes the difference between output and HEVC baseline. Positive indicates performance improvement. The black font is the best.
Result

Visual Quality

Figure 5: Visual Quality Comparisons in test 2. The left column is original frame. The middle column is frame compressed by HEVC. The right column is output of our model.


Reference
Thanks !