Perceptual Quality Assessment of Internet Videos

Jiahua Xu, Alibaba Group, USTC

Jing Li*, Alibaba Group

XingGuang Zhou, Alibaba Group

Wei Zhou, USTC

Baichao Wang, Alibaba Group

Zhibo Chen, USTC
Internet Videos

**UGC**
- Captured, edited, uploaded by users
- Diverse contents and uncontrolled quality

**PGC**
- Professional device and photographer
- Well-designed contents and guaranteed quality

**OGC**
- e.g. Hollywood production
Diverse contents in one website

• **A general quality assessment solution**
  Input: any videos
  Output: perceptual quality score

• **No reference in most cases**
  Input: only the videos under test
  Output: perceptual quality score for this video
Challenges

• **Data**
  - How to collect the subjective score → reliable label
  - How many data should we collect → for DL model
  - How to run the subjective test with limited budget → reality
  - How to remove outliers when we have no ground truth → reliable label

• **Model**
  How to deal with the human perception in such a complicated case?
  Universal
Database: Youku-V1K

- **Data sampling strategy**
  - Full coverage
  - Uniform
  - Small&Valid

  **Coarse sampling**
  - Randomly sampling 10K videos from Youku
  - Uniformly sampling 3K videos from above by content category and resolution

  **Fine sampling**
  - Calculating quality factors, including spatial activity, temporal activity, blockiness, blurriness, brightness, contrast, flickering, colorfulness, etc.
  - Sampling videos to make each factor as uniformly distributed as possible
Database: Youku-V1K

Resolution distribution

- 540p: 27%
- 720p: 43%
- 1080p: 30%

Content category distribution

- Drama: 14%
- Vlog: 14%
- Gaming: 5%
- Shows: 4%
- Cartoon: 4%
- Sports: 4%
- Music: 7%
- News: 4%
- Film: 12%
- Entertainment: 9%
- Scenery: 9%
- Education: 8%
- Others: 7%
### Database: Youku-V1K

- **1072 videos**
- **540p – 1080p**
- **13 content categories**
- **UGC + PGC + OGC**

#### Video Quality and Distortion Distribution

<table>
<thead>
<tr>
<th>Source</th>
<th>Videos</th>
<th>Video Length</th>
<th>Resolution</th>
<th>Distortion Type</th>
<th>Subjective Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Youku</td>
<td>1072</td>
<td>10s</td>
<td>1080p</td>
<td>authentic (UGC+PGC+OGC)</td>
<td>Crowdsourcing</td>
</tr>
</tbody>
</table>
Subjective experiment

• Absolute Categorical Rating (ACR) method

• Crowdsourcing
  • 300+ naïve observers
  • Aged from 18-49 years old
  • Approximately 100 votings /observer, avoiding visual fatigue
  • 22000+ labeling
  • 15+ labels/video
  • Videos are randomly presented
  • Viewed on PC, viewing distance uncontrolled
Data Cleaning

• Adopted a probabilistic graphic Model\textsuperscript{[1]} for data cleaning

The proposed objective quality model is designed to capture the spatial relations using a GCN (Graph Convolutional Network). This approach is integrated into a frame-level model, which includes feature extraction layers such as ResNet-18. The model also incorporates an attention module for channel attention and a similarity matrix normalization component. The optical flow module is another key component that processes the flow data.
The proposed objective quality model

Attention:

to enhance the features for discriminative Channels and salient regions
The proposed objective quality model

Optical Flow: to capture motion information from adjacent frames
The proposed objective quality model uses a bi-directional LSTM to capture long-term inter-frames relations, i.e., quality fluctuations.
Experimental results:

<table>
<thead>
<tr>
<th>Quality metrics</th>
<th>Video databases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SROCC</td>
</tr>
<tr>
<td>NIQE</td>
<td>0.5782(±0.0112)</td>
</tr>
<tr>
<td>IJNIQE</td>
<td>0.4427(±0.0121)</td>
</tr>
<tr>
<td>VIIDEO</td>
<td>0.4210(±0.0124)</td>
</tr>
<tr>
<td>BRISQUE</td>
<td>0.7804(±0.0268)</td>
</tr>
<tr>
<td>GM-LOG</td>
<td>0.7930(±0.0241)</td>
</tr>
<tr>
<td>HIGRADE</td>
<td>0.8486(±0.0170)</td>
</tr>
<tr>
<td>FRIQUEE</td>
<td>0.8512(±0.0182)</td>
</tr>
<tr>
<td>CORINA</td>
<td>0.8464(±0.0176)</td>
</tr>
<tr>
<td>HOSA</td>
<td>0.8480(±0.0144)</td>
</tr>
<tr>
<td>VGG-19</td>
<td>0.8647(±0.0180)</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.8791(±0.0157)</td>
</tr>
<tr>
<td>V-BLIINDS</td>
<td>0.7822(±0.0245)</td>
</tr>
<tr>
<td>TLVQM</td>
<td>0.7832(±0.0237)</td>
</tr>
<tr>
<td>VIDEVAL</td>
<td>0.8294(±0.0183)</td>
</tr>
<tr>
<td>STDAM</td>
<td><strong>0.9141(±0.0089)</strong></td>
</tr>
</tbody>
</table>
Experimental results:

<table>
<thead>
<tr>
<th>Quality metrics</th>
<th>Youku-V1K</th>
<th>KoNViD-1k</th>
<th>LIVE-VQC</th>
<th>YouTube-UGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIQE</td>
<td>0.6046(±0.0097)</td>
<td>0.5530(±0.0337)</td>
<td>0.6286(±0.0512)</td>
<td>0.2776(±0.0431)</td>
</tr>
<tr>
<td>ILNIQE</td>
<td>0.4685(±0.0110)</td>
<td>0.5400(±0.0337)</td>
<td>0.5437(±0.0717)</td>
<td>0.3302(±0.0579)</td>
</tr>
<tr>
<td>VIEDEO</td>
<td>0.4148(±0.0128)</td>
<td>0.3002(±0.0539)</td>
<td>0.2146(±0.0903)</td>
<td>0.1534(±0.0498)</td>
</tr>
<tr>
<td>BRISQUE</td>
<td>0.7801(±0.0278)</td>
<td>0.6576(±0.0342)</td>
<td>0.6380(±0.0632)</td>
<td>0.3952(±0.0486)</td>
</tr>
<tr>
<td>GM-LOG</td>
<td>0.7958(±0.0545)</td>
<td>0.6636(±0.0315)</td>
<td>0.6212(±0.0636)</td>
<td>0.3920(±0.0594)</td>
</tr>
<tr>
<td>HIGRADE</td>
<td>0.8507(±0.0166)</td>
<td>0.7269(±0.0287)</td>
<td>0.6332(±0.0652)</td>
<td>0.7216(±0.0334)</td>
</tr>
<tr>
<td>FRIQUEE</td>
<td>0.8508(±0.0185)</td>
<td>0.7482(±0.0257)</td>
<td>0.7000(±0.0587)</td>
<td>0.7571(±0.0324)</td>
</tr>
<tr>
<td>CORINA</td>
<td>0.8479(±0.0188)</td>
<td>0.7135(±0.0236)</td>
<td>0.7183(±0.0420)</td>
<td>0.6057(±0.0399)</td>
</tr>
<tr>
<td>HOSA</td>
<td>0.8485(±0.0144)</td>
<td>0.7664(±0.0207)</td>
<td>0.7414(±0.0410)</td>
<td>0.6047(±0.0347)</td>
</tr>
<tr>
<td>VGG-19</td>
<td>0.8704(±0.0156)</td>
<td>0.7845(±0.0246)</td>
<td>0.7160(±0.0481)</td>
<td>0.6997(±0.0281)</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.8821(±0.0154)</td>
<td>0.8104(±0.0229)</td>
<td>0.7205(±0.0434)</td>
<td>0.7097(±0.0276)</td>
</tr>
<tr>
<td>V-BLIINDS</td>
<td>0.7844(±0.0249)</td>
<td>0.7037(±0.0301)</td>
<td>0.7178(±0.0500)</td>
<td>0.5551(±0.0465)</td>
</tr>
<tr>
<td>TLVQM</td>
<td>0.7849(±0.0243)</td>
<td>0.7688(±0.0238)</td>
<td>0.8025(±0.0360)</td>
<td>0.6590(±0.0302)</td>
</tr>
<tr>
<td>VIDEVAL</td>
<td>0.8304(±0.0187)</td>
<td>0.7803(±0.0233)</td>
<td>0.7514(±0.0420)</td>
<td>0.7733(±0.0257)</td>
</tr>
<tr>
<td>STDAM</td>
<td><strong>0.9120(±0.0074)</strong></td>
<td><strong>0.8415(±0.0173)</strong></td>
<td><strong>0.8204(±0.0342)</strong></td>
<td><strong>0.8297(±0.0279)</strong></td>
</tr>
</tbody>
</table>
Applications

The proposed model has been widely used at Youku

- Quality score as a ranking factor in recommendation systems
- Low-quality filtering in searching systems
- Low-quality filtering when users uploading their videos
- Quality enhancement indicators
Thank you!