ACM multimedia

Chengdu, China OCT 20-24 2021

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

阿里巴巴 Alibaba.com

中国科学技术大学 University of Science and Technology of China

Internet Videos

UGC

- Captured, edited, uploaded by users
- Diverse contents and un-controlled quality

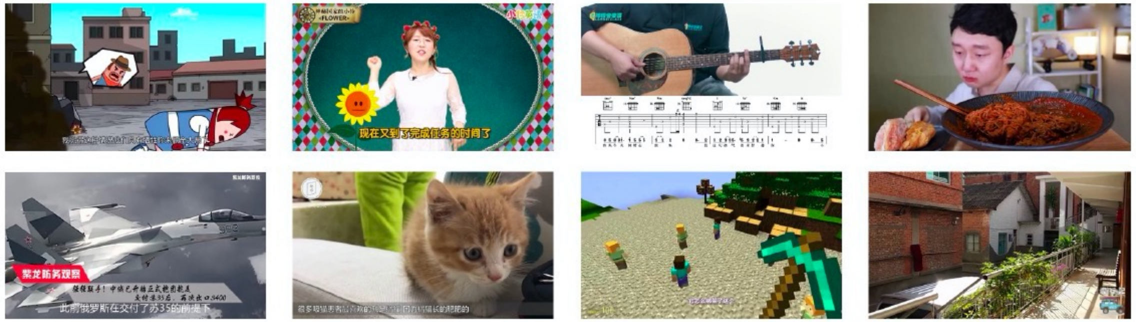
PGC

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

OGC

- e.g. Hollywood production

User generated content (UGC) videos



Professionally generated content (PGC) videos

Occupationally generated content (OGC) videos

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

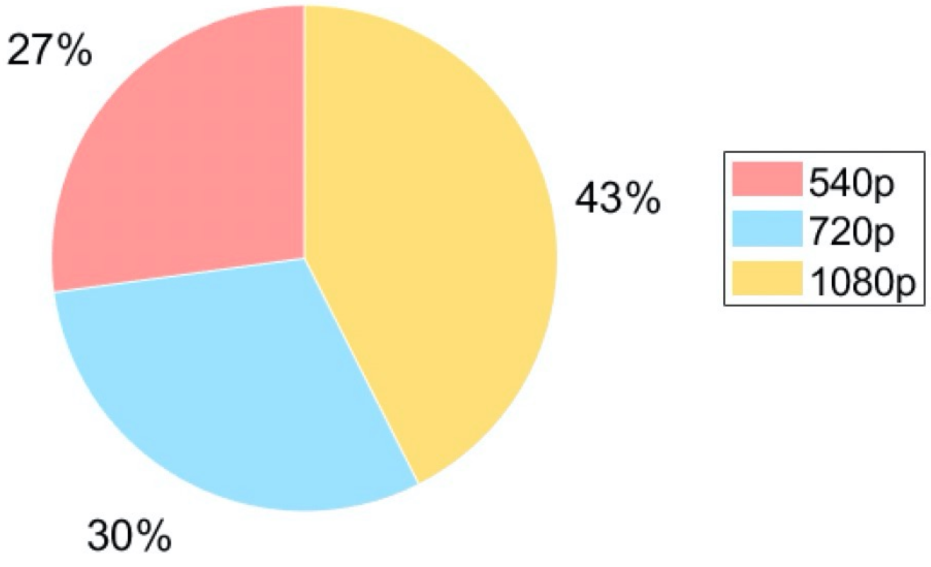
- Randomly sampling 10K videos from Youku

- Uniformly sampling 3K videos from above by content category and resolution

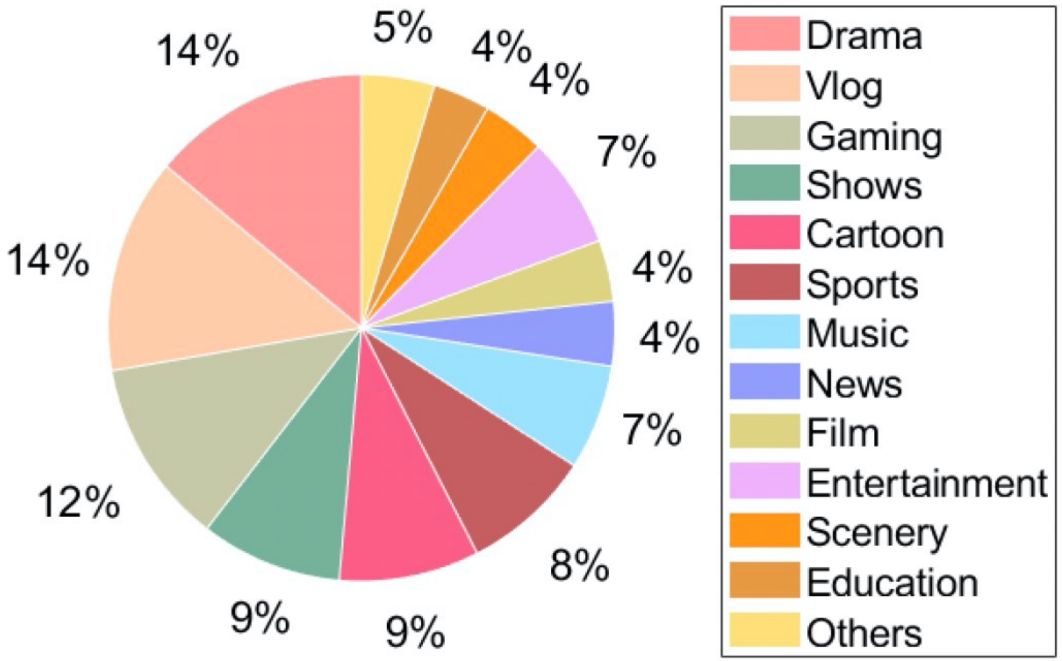
- Calculating quality factors, including spatial activity, temporal activity, blockiness, blurriness, brightness, contrast, flickering, colorfulness, etc.

- Sampling videos to make each factor as uniformly distributed as possible

Database: Youku-V1K



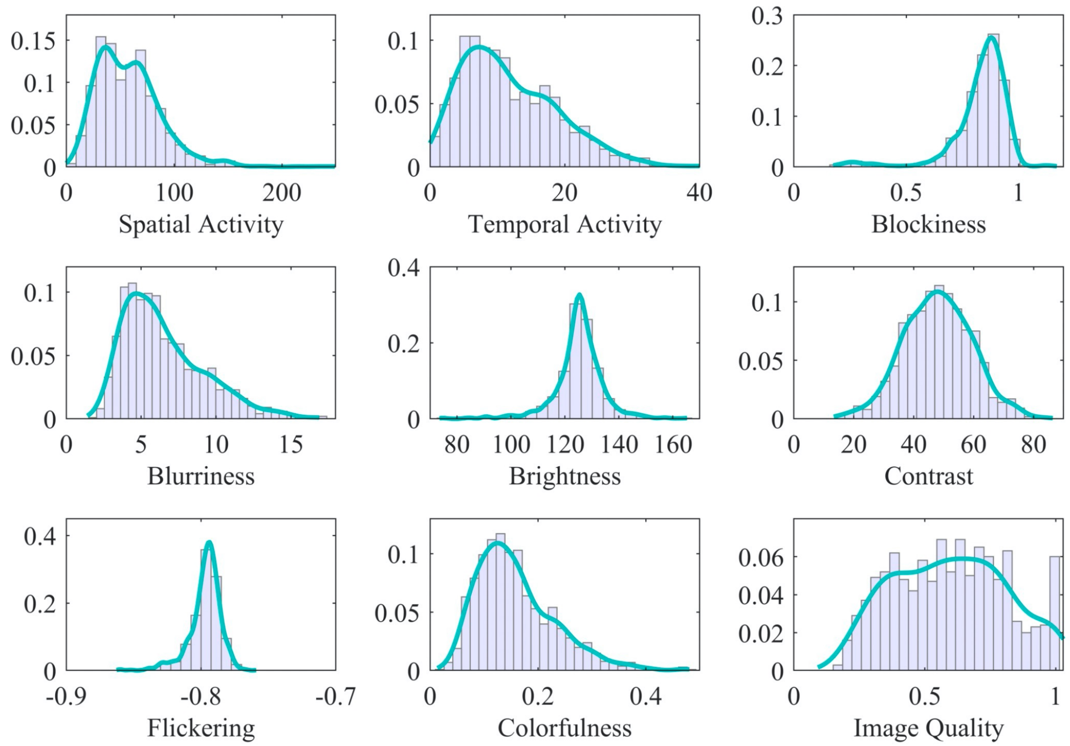
Resolution distribution



Content category distribution

Database: Youku-V1K

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



	Databases	Source	# of videos(Ref/Dis)	Video length	Resolution	Distortion type	Subjective environment
With high-quality reference	IRCCyN/IVC 1080i [35]	High-quality reference	24/168	9-12s	1080p	synthetic	Laboratory
	LIVE [40]	High-quality reference	10/150	8-10s	768x432	synthetic	Laboratory
	CSIQ [45]	High-quality reference	12/216	10s	832x480	synthetic	Laboratory
Without high-quality reference	CVD2014 [34]	Captured	-/234	10-25s	480p, 720p	authentic	Laboratory
	LIVE-Qualcomm [12]	Captured	-/208	15s	1080p	authentic	Laboratory
	LIVE-VQC [42]	Captured	-/585	10s	480p-1080p	authentic	Crowdsourcing
	KoNViD-1k [17]	Flicker	-/1200	8s	540p	authentic (UGC)	Crowdsourcing
	YouTube-UGC [44]	YouTube	-/1380	20s	360p-2160p	authentic (UGC)	Crowdsourcing
	Youku-V1K	Youku	-/1072	10s	1080p	authentic (UGC+PGC+OGC)	Crowdsourcing

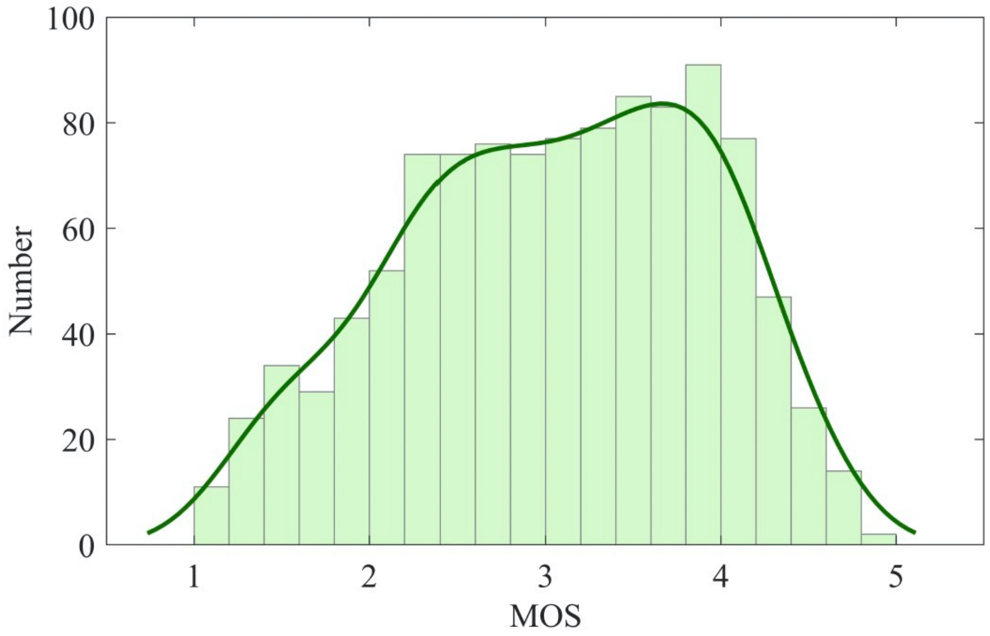
Subjective experiment

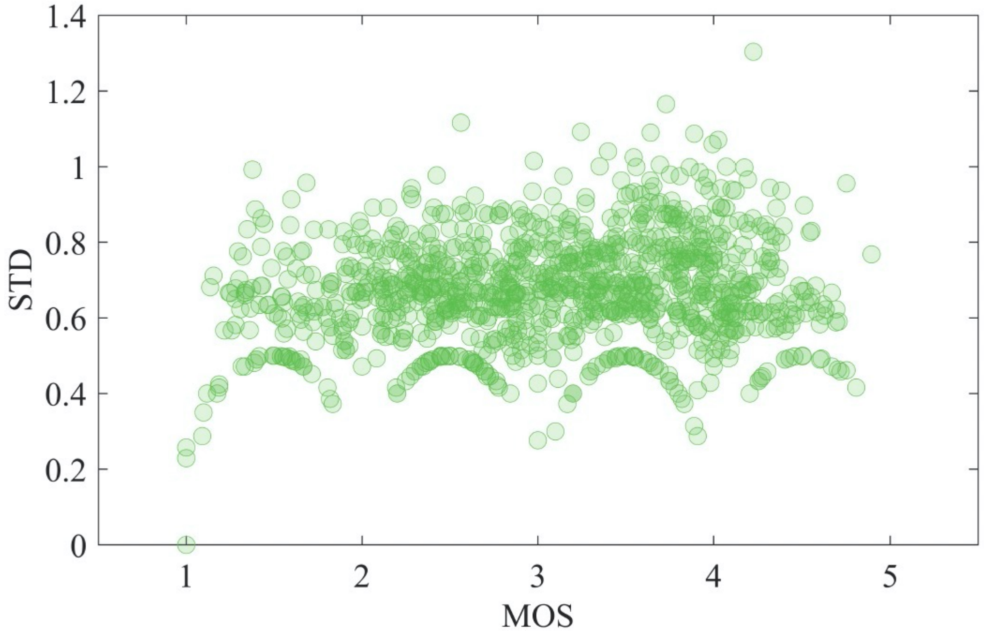
- Absolute Categorical Rating (ACR) method

- 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[1] for data cleaning



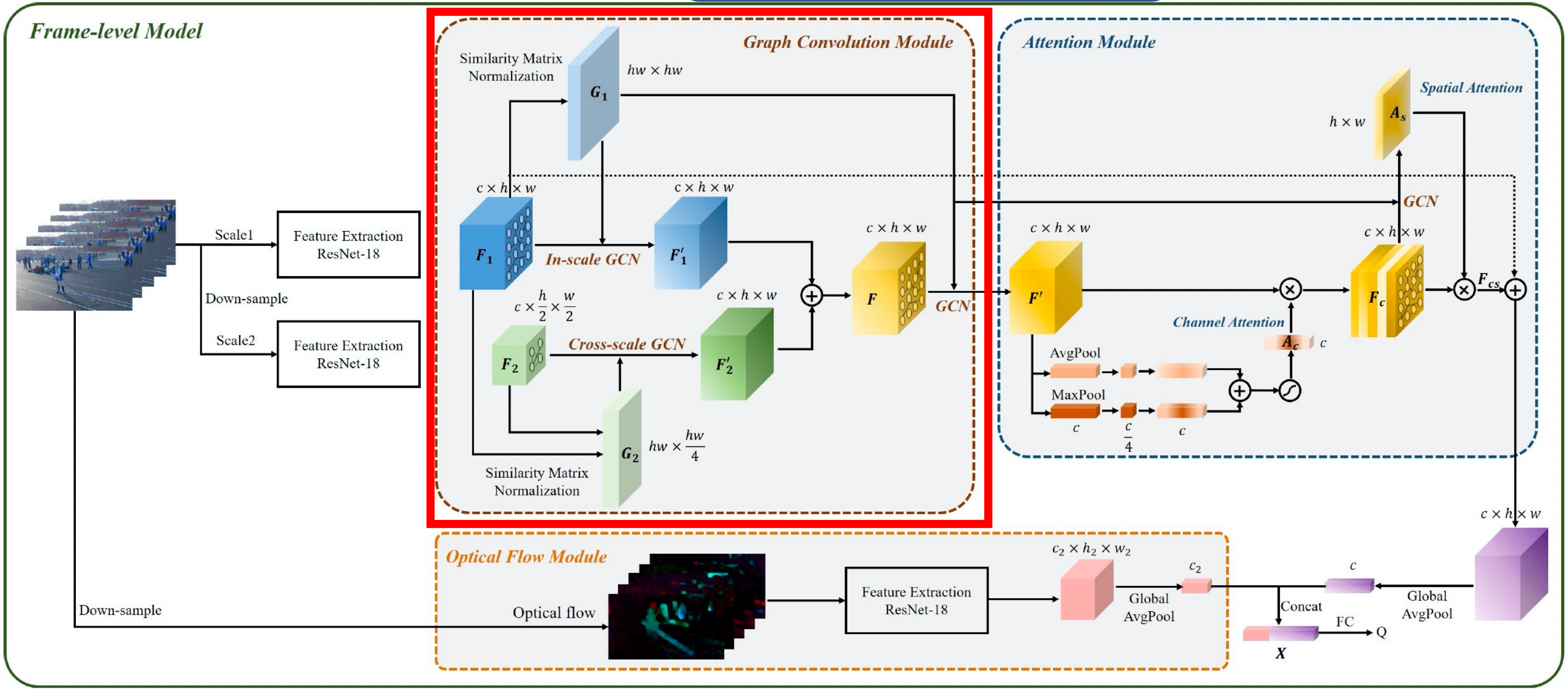


[1]J Li, S Ling, J Wang, Z Li, and P Callet. *A probabilistic graphical model for analyzing the subjective visual quality assessment data from crowdsourcing.* In Proceedings of ACM MM, 2020.

The proposed obje odel

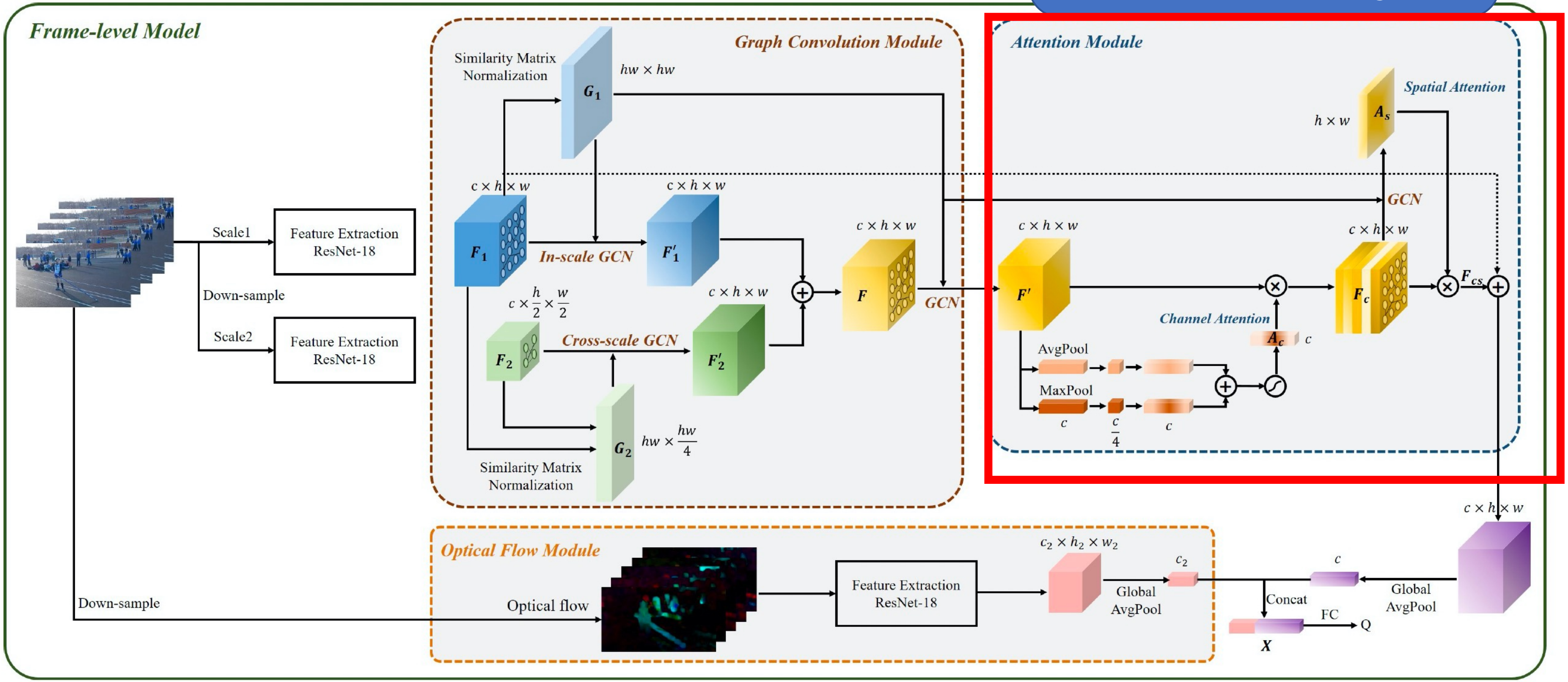
GCN

To capture the spatial relations

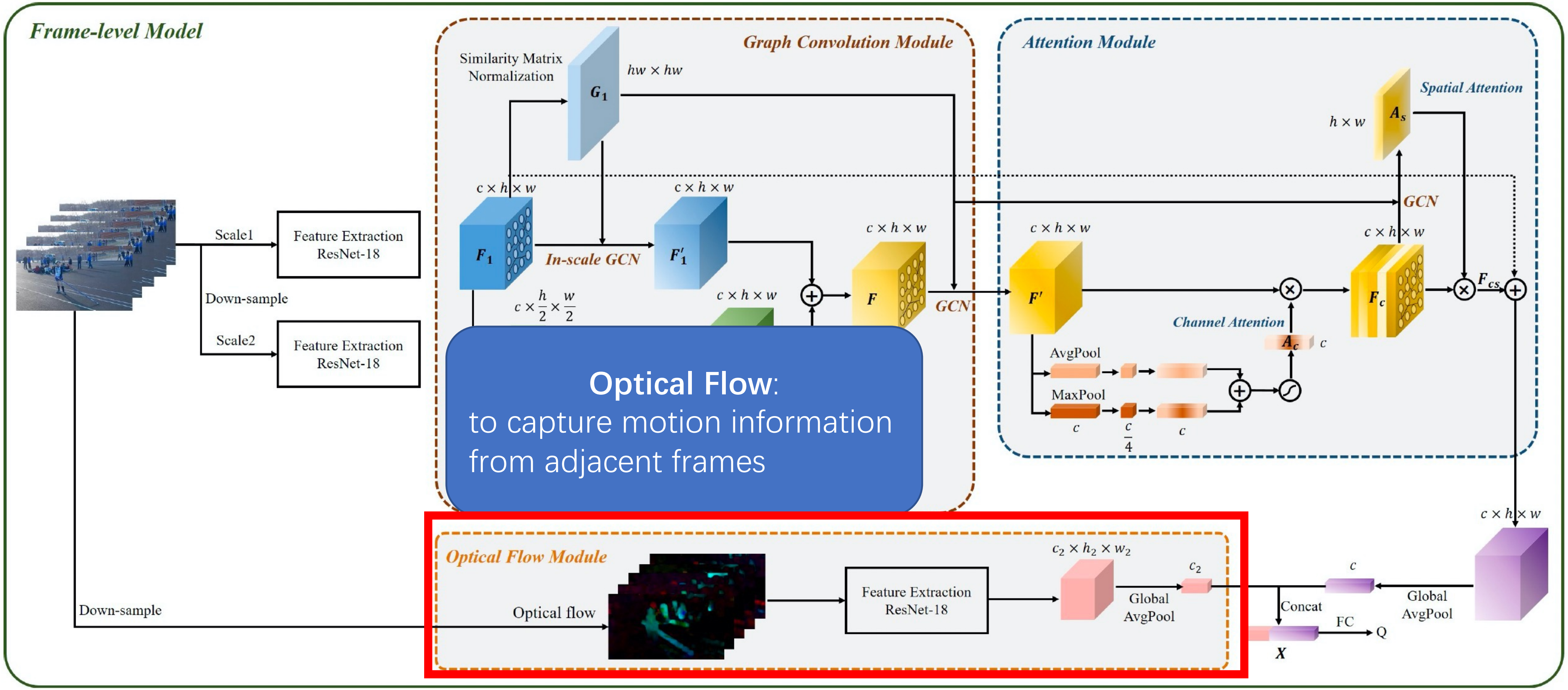


The proposed objective quality

Attention:
to enhance the features for discriminative
Channels and salient regions



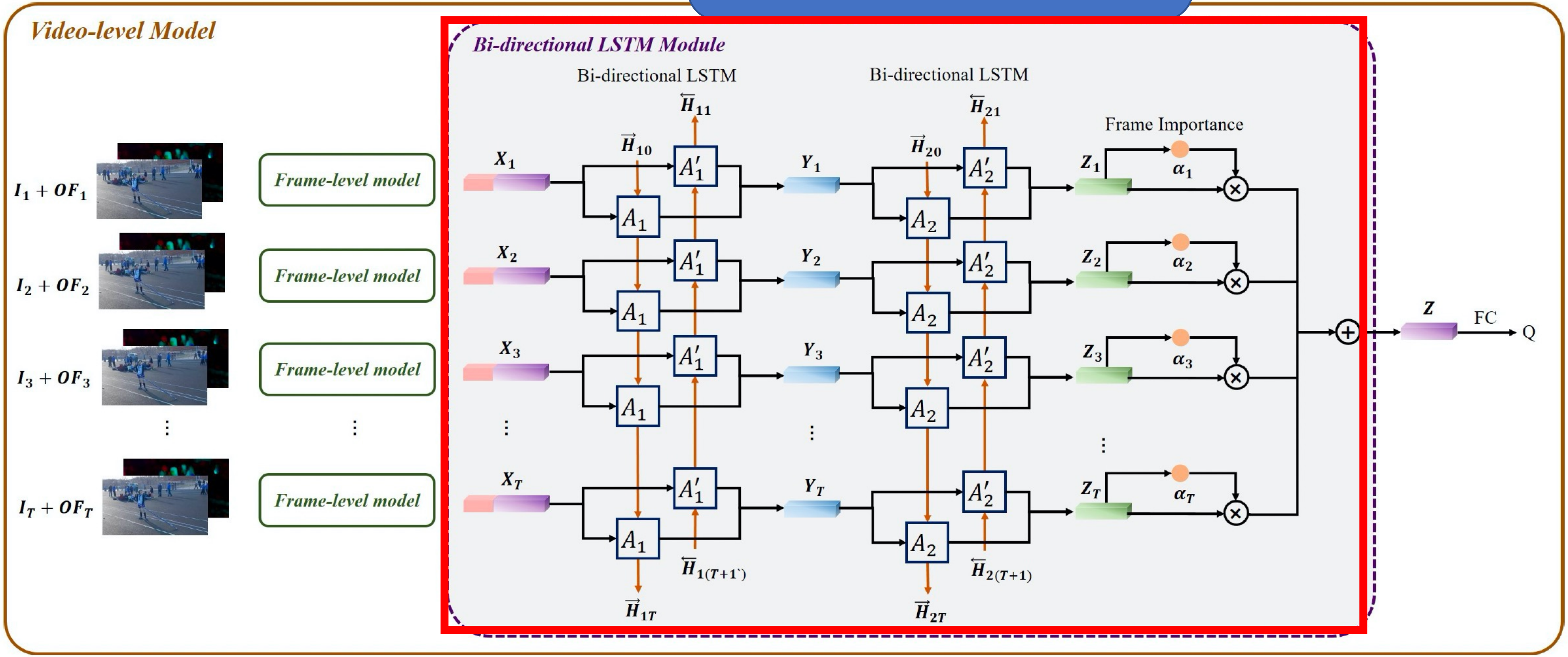
The proposed objective quality model



Optical Flow:
to capture motion information from adjacent frames

The proposed obje… del

Bi-directional LSTM:
to capture long-term inter-frames relations, i.e., quality fluctuations



Experimental results:

Video databases

Quality metrics

SROCC	Youku-V1K	KoNViD-1k	LIVE-VQC	YouTube-UGC
NIQE	0.5782(±0.0112)	0.5417(±0.0347)	0.5957(±0.0571)	0.2379(±0.0487)
ILNIQE	0.4427(±0.0121)	0.5264(±0.0294)	0.5037(±0.0712)	0.2918(±0.0502)
VIIDEO	0.4210(±0.0124)	0.2988(±0.0561)	0.0332(±0.0856)	0.0580(±0.0536)
BRISQUE	0.7804(±0.0268)	0.6567(±0.0351)	0.5929(±0.0681)	0.3820(±0.0519)
GM-LOG	0.7930(±0.0241)	0.6578(±0.0324)	0.5881(±0.0683)	0.3678(±0.0589)
HIGRADE	0.8486(±0.0170)	0.7206(±0.0302)	0.6103(±0.0680)	0.7376(±0.0338)
FRIQUEE	0.8512(±0.0182)	0.7472(±0.0263)	0.6579(±0.0536)	0.7652(±0.0301)
CORINA	0.8464(±0.0176)	0.7169(±0.0245)	0.6719(±0.0473)	0.5972(±0.0413)
HOSA	0.8480(±0.0144)	0.7654(±0.0224)	0.6873(±0.0462)	0.6025(±0.0344)
VGG-19	0.8647(±0.0180)	0.7741(±0.0288)	0.6568(±0.0536)	0.7025(±0.0281)
ResNet-50	0.8791(±0.0157)	0.8018(±0.0255)	0.6636(±0.0511)	0.7183(±0.0281)
V-BLIINDS	0.7822(±0.0245)	0.7101(±0.0314)	0.6939(±0.0502)	0.5590(±0.0496)
TLVQM	0.7832(±0.0237)	0.7729(±0.0242)	**0.7988(±0.0365)**	0.6693(±0.0306)
VIDEVAL	0.8294(±0.0183)	0.7832(±0.0216)	0.7522(±0.0390)	0.7787(±0.0254)
STDAM	**0.9141(±0.0089)**	**0.8448(±0.0189)**	0.7931(±0.0340)	**0.8341(±0.0306)**

Experimental results:

Quality metrics	Video databases			
PLCC	Youku-V1K	KoNViD-1k	LIVE-VQC	YouTube-UGC
NIQE	0.6046(±0.0097)	0.5530(±0.0337)	0.6286(±0.0512)	0.2776(±0.0431)
ILNIQE	0.4685(±0.0110)	0.5400(±0.0337)	0.5437(±0.0717)	0.3302(±0.0579)
VIIDEO	0.4148(±0.0128)	0.3002(±0.0539)	0.2146(±0.0903)	0.1534(±0.0498)
BRISQUE	0.7801(±0.0278)	0.6576(±0.0342)	0.6380(±0.0632)	0.3952(±0.0486)
GM-LOG	0.7958(±0.0545)	0.6636(±0.0315)	0.6212(±0.0636)	0.3920(±0.0594)
HIGRADE	0.8507(±0.0166)	0.7269(±0.0287)	0.6332(±0.0652)	0.7216(±0.0334)
FRIQUEE	0.8508(±0.0185)	0.7482(±0.0257)	0.7000(±0.0587)	07571(±0.0324)
CORINA	0.8479(±0.0188)	0.7135(±0.0236)	0.7183(±0.0420)	0.6057(±0.0399)
HOSA	0.8485(±0.0144)	0.7664(±0.0207)	0.7414(±0.0410)	0.6047(±0.0347)
VGG-19	0.8704(±0.0156)	0.7845(±0.0246)	0.7160(±0.0481)	0.6997(±0.0281)
ResNet-50	<u>0.8821(±0.0154)</u>	<u>0.8104(±0.0229)</u>	0.7205(±0.0434)	0.7097(±0.0276)
V-BLIINDS	0.7844(±0.0249)	0.7037(±0.0301)	0.7178(±0.0500)	0.5551(±0.0465)
TLVQM	0.7849(±0.0243)	0.7688(±0.0238)	<u>0.8025(±0.0360)</u>	0.6590(±0.0302)
VIDEVAL	0.8304(±0.0187)	0.7803(±0.0233)	0.7514(±0.0420)	<u>0.7733(±0.0257)</u>
STDAM	**0.9120(±0.0074)**	**0.8415(±0.0173)**	**0.8204(±0.0342)**	**0.8297(±0.0279)**

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

ACM multimedia

Chengdu, China OCT 20-24 2021

Thank you!

阿里巴巴 Alibaba.com

中国科学技术大学 University of Science and Technology of China