VQA on CG Videos

Analyzing Video Quality Assessment Methods' on Computer Graphic Videos

Data Science, University of Twente
Ferhat Ege Darici
Supervised by Estefanía Talavera Martínez

BACKGROUND & RELATED WORK

VIDEO EDITING

Manual assessment of video quality is hard.

ONLY TECHNICAL

VQAs usually check for blurs and distortions.

SUBJECTIVE

Content & composition preference may differ.

COMPUTER GRAPHICS

As a novelty, computer graphics videos are assessed.

01 **STUDY**
RQs and Methodology

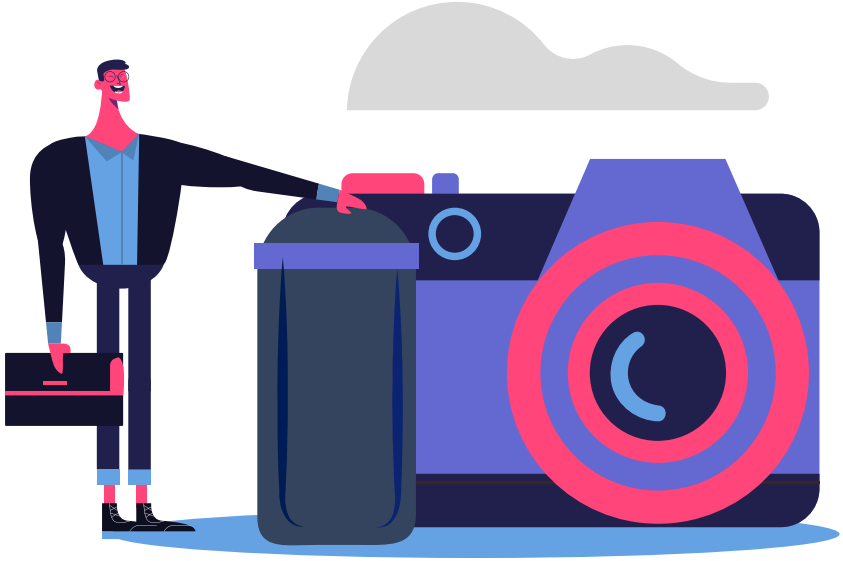
02 **VQA METHODS**
and Datasets

03 **RESULTS**
Graphs and Tables

04 **CONCLUSION**
and Discussion

01 STUDY

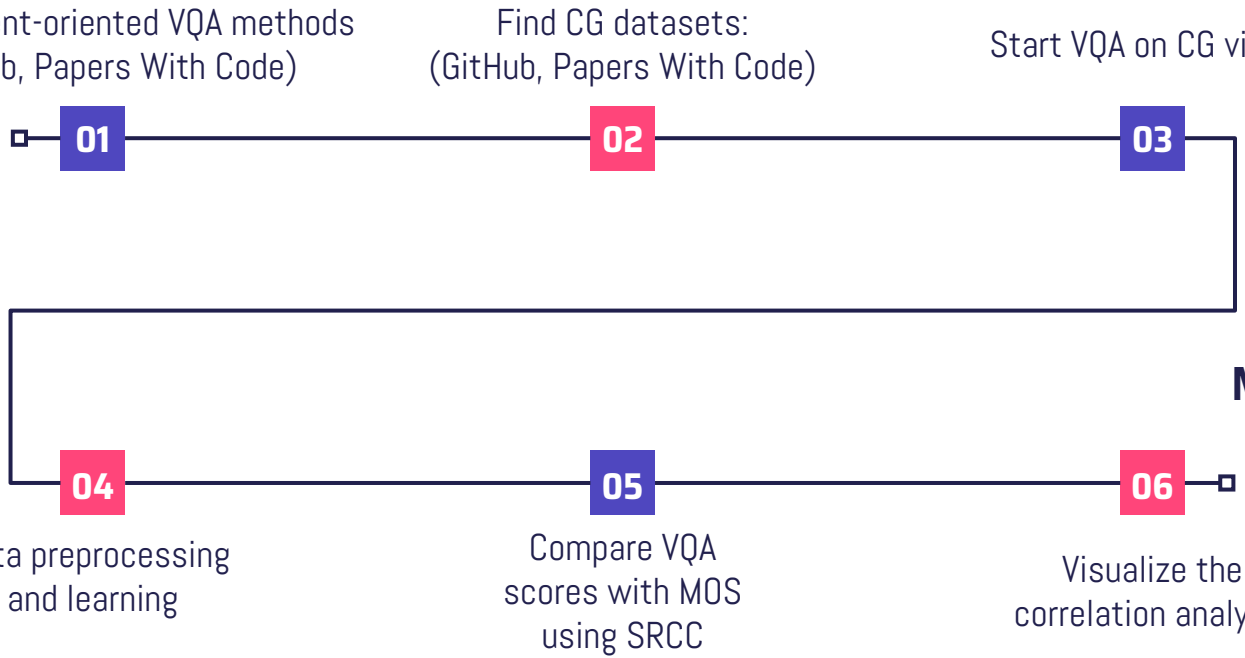
Research Question and Plan

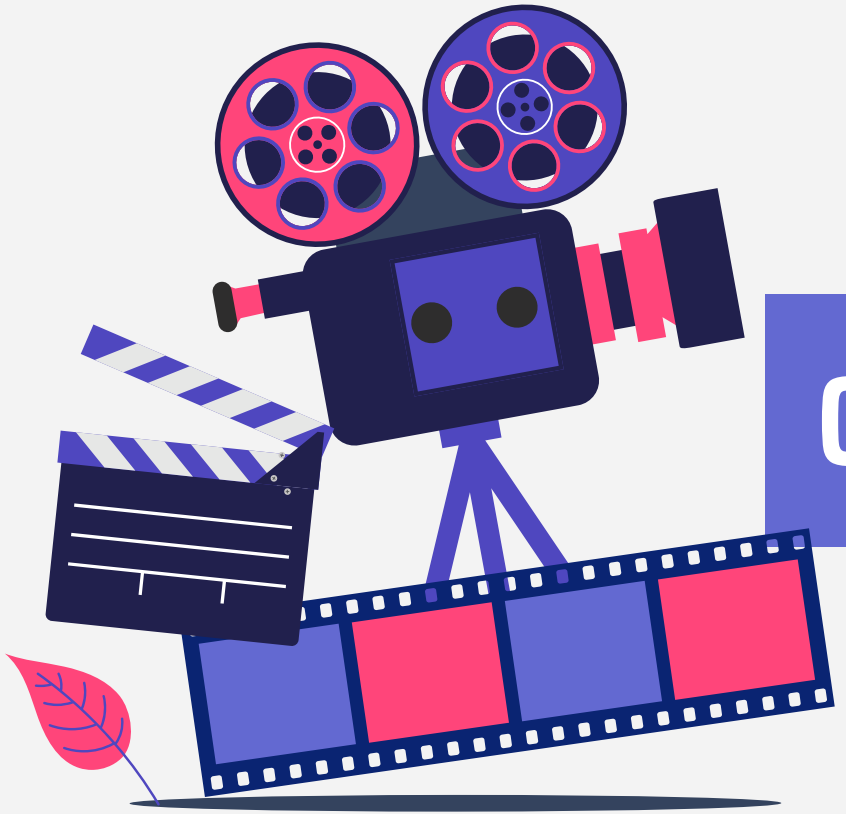


Research Question

How is the performance of video quality assessment (VQA) methods on computer graphic (CG) videos (e.g. animation, gaming) compared to mean opinion scores (MOS) on videos?

METHODOLOGY





02

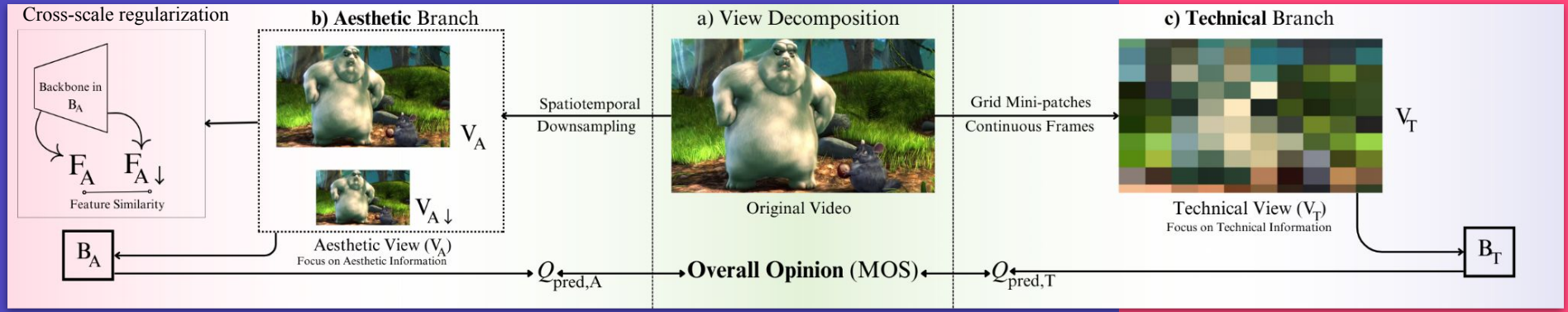
VQA METHODS

and Datasets

Chosen VQA and Why?

The main VQA method I choose to work with is the **Disentangled Objective Video Quality Evaluator (DOVER)**. Because:

- Uses View Decomposition strategy
  - **Separate the video in two views**: Aesthetic view and Technical view



DATASET

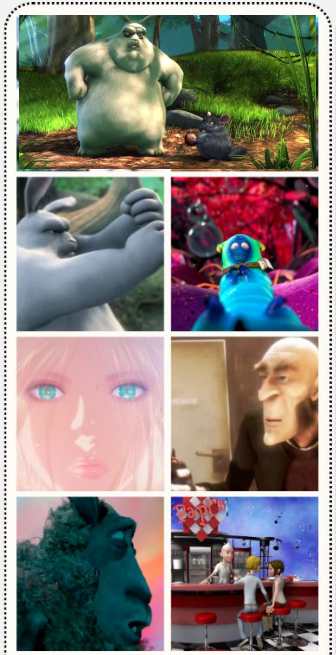
CG Animation Dataset

27 High quality reference videos
262 Distorted videos:

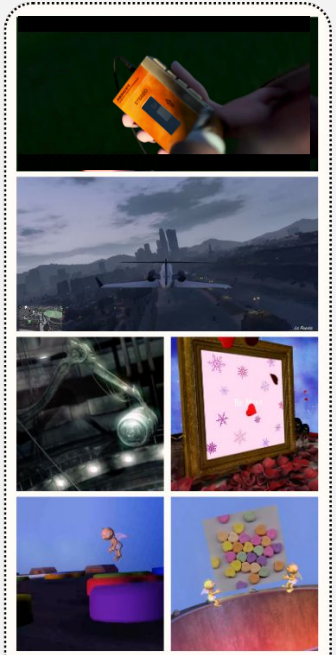
- AVC/H.264 compression
- HEVC/H.265 compression
- MPEG-2 compression

Provides:

- Animation videos
- Gaming videos
- MOS for each video



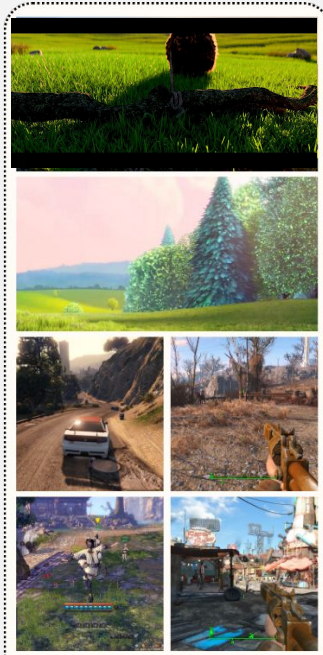
Character and Face (CF)



Items in front of Simple Background (ISB)

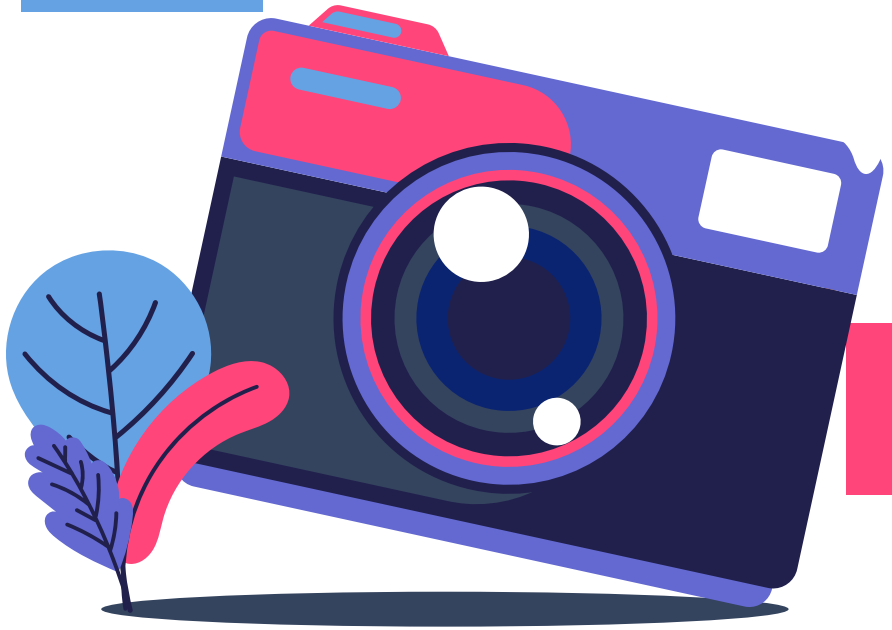
Gorgeous Special Effects (GSE)

Multiplayer Online Battle Arena (MOBA)



Scenery and Architecture (SA)

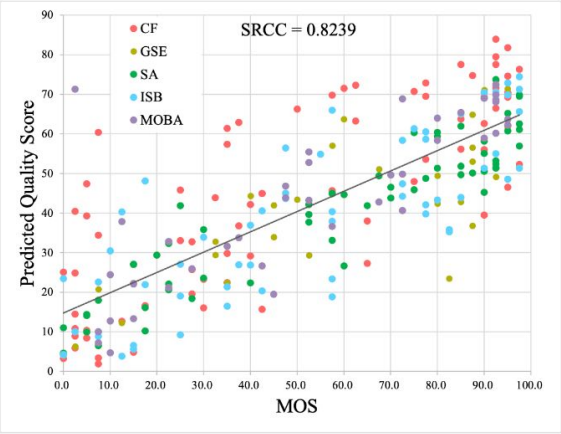
03

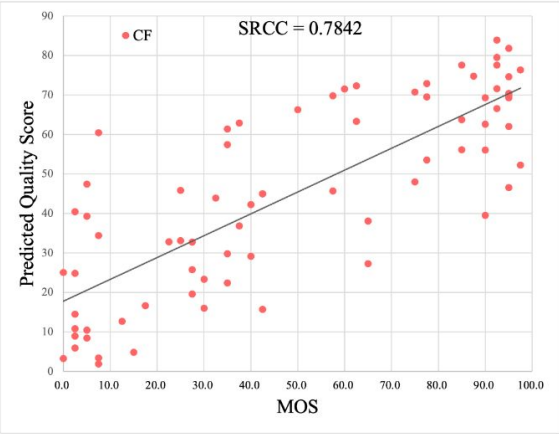


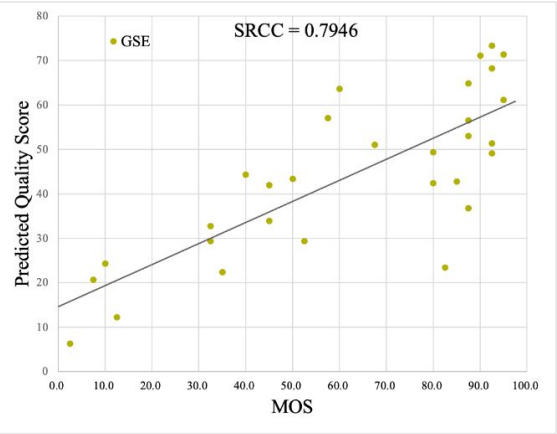
RESULTS

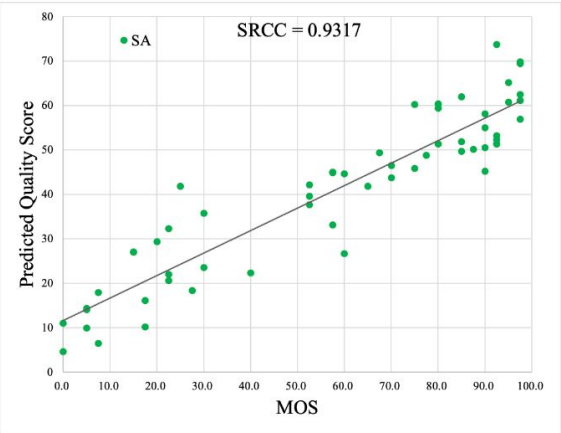
Comparison and Analysis

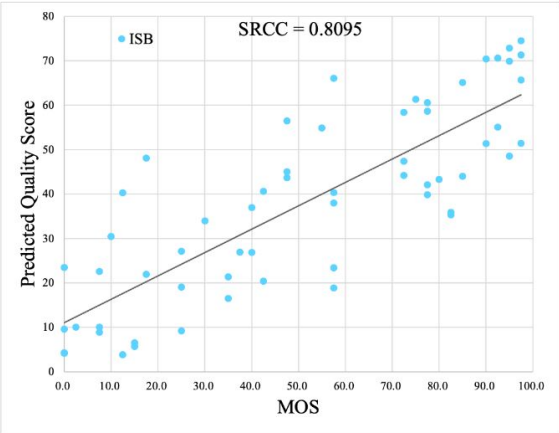
SCATTER PLOT ANALYSIS

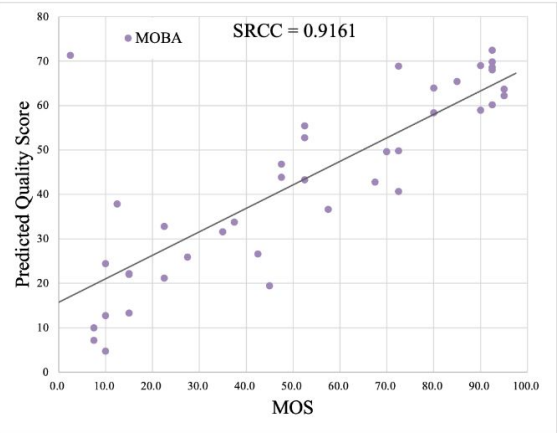












Spearman's Rank Correlation: 82% (High)

COMPARISON TABLE ANALYSIS

Method	CF	GSE	SA	ISB	MOBA	Overall
PSNR [6]	0.83021	0.5599	-0.14103	0.6087	0.4414	0.31273
SSIM [23]	0.75001	0.49208	-0.01444	0.47698	0.66604	0.31056
VMAF [47]	**0.87226**	0.79211	0.4292	0.80942	0.73926	0.57745
DOVER [37]	0.7842	**0.7946**	**0.9317**	**0.8095**	**0.9161**	**0.8239**

PSNR: Peak Signal-to-Noise Ratio
SSIM: Structural Similarity

Check pixel-wise differences
- Good at capturing repeating patterns, well-defined structures
- Not good at capturing details and textures

Correlation: **~31% (Not good)**

VMAF: Video Multi-method Assessment Fusion

Checks spatial, temporal and motion information
- Not good at capturing semantics in scenes

Correlation: **~58% (Fair/Good)**

DOVER: Disentangled Objective Video Quality Evaluator

Checks both technical and aesthetic aspects
- Good at capturing semantics in scenes

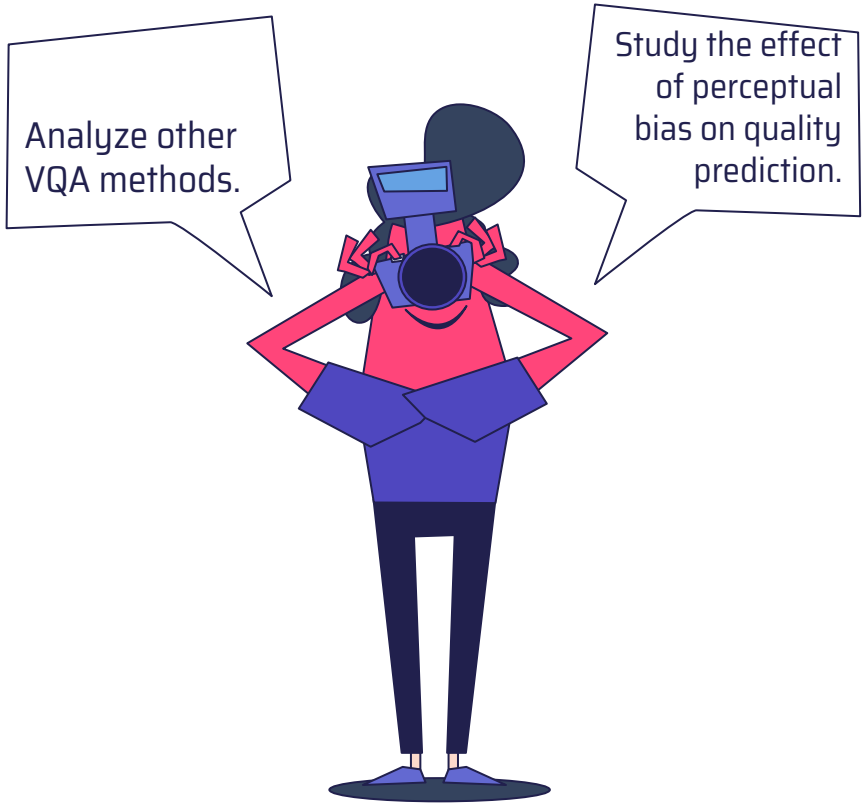
Correlation: **~82% (Very good)**

04

CONCLUSION

And Discussion

DISCUSSION


Analyze other VQA methods.
Study the effect of perceptual bias on quality prediction.

Research Question

How is the performance of VQA methods on CG videos (e.g. animation, gaming) compared to mean opinion scores (MOS) on videos?

Content-aware VQA has **better** correlation with **human perception**.

Questions

Thanks!

References

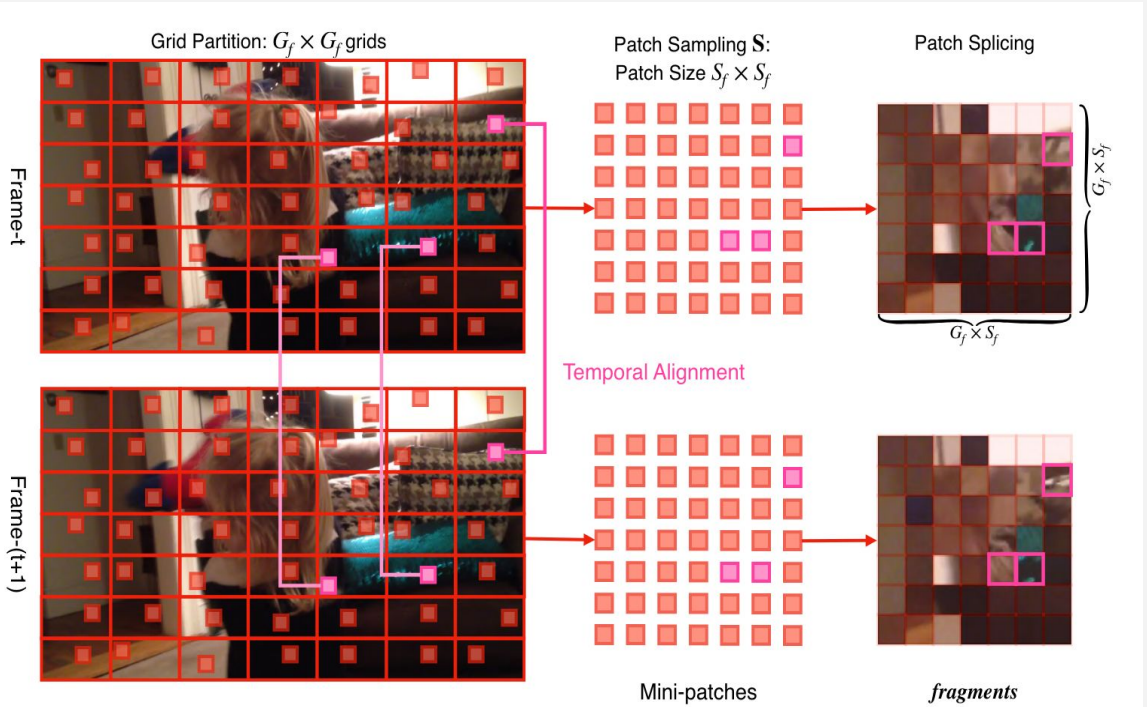
[1]Haoning Wu,Erl iZhang,Liang Liao,Chaofeng Chen,Jingwen Hou,AnnanWang, Wenxiu Sun, Qiong Yan, and Weisi Lin. 2023. Exploring Video Quality Assessment on User Generated Contents from Aesthetic and Technical Perspectives. http: //arxiv.org/abs/2211.04894 arXiv:2211.04894

[2]Haoning Wu,Chaofeng Chen,Jingwen Hou,Liang Liao,Annan Wang,Wenxiu Sun, Qiong Yan, and Weisi Lin. 2022. FAST-VQA: Efficient End-to-end Video Quality Assessment with Fragment Sampling. https://arxiv-org.ezproxy2.utwente. nl/abs/2207.02595v1

- The backbone of the technical branch is Video Swin Transformer Tiny with Gated Relative Position Biases (GRPB).
  - a pure-transformer backbone architecture for video recognition that is found to surpass the factorized models in efficiency.
- And the aesthetic backbone is Conv-next Tiny pre-trained with AVA which is an aesthetic assessment dataset.

Out-of-scope Information

[2]



[2]

Fig. 4: The pipeline for sampling ***fragments*** with Grid Mini-patch Sampling (GMS), including grid partition, patch sampling, patch splicing, and temporal alignment. After GMS, the ***fragments*** are fed into the FANet (Fig. 5).

[2]

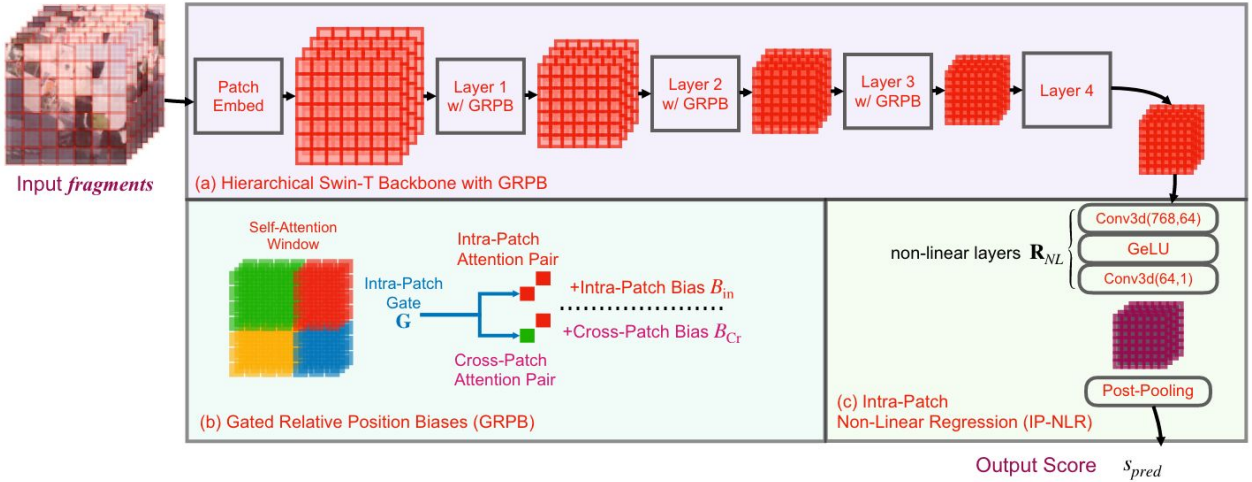


Fig. 5: The overall framework for FANet, including the Gated Relative Position Biases (GRPB) and Intra-Patch Non-Linear Regression (IP-NLR) modules. The input ***fragments*** come from Grid Mini-patch Sampling (Fig. 4).

SRCC

- Performance calculated by Spearman correlation coefficient
  - Calculate difference between VQA score and MOS for each video
  - Square the differences
  - Sum the squared differences
  - Calculate the correlation coefficient:

    $\rho = 1 - (6 * \Sigma d^2) / (n * (n^2 - 1))$

    where ρ is the Spearman correlation coefficient, d is the difference in ranks, and n is the number of data points.