NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

SCSE23-0760 Video Quality Assessment Modelling

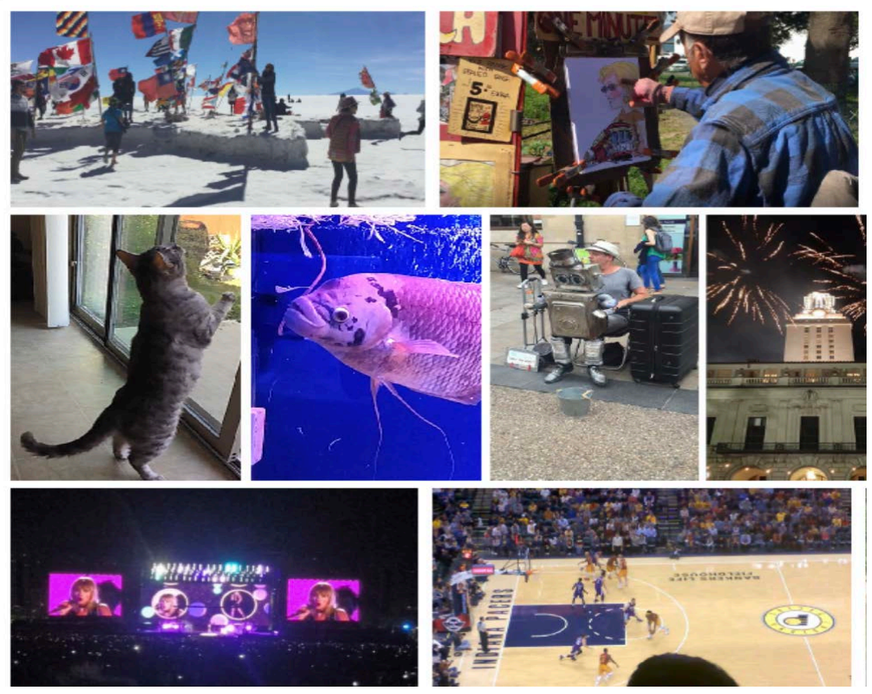
Student: Wang Chuhan Supervisor: Prof Lin Weisi

Motivation

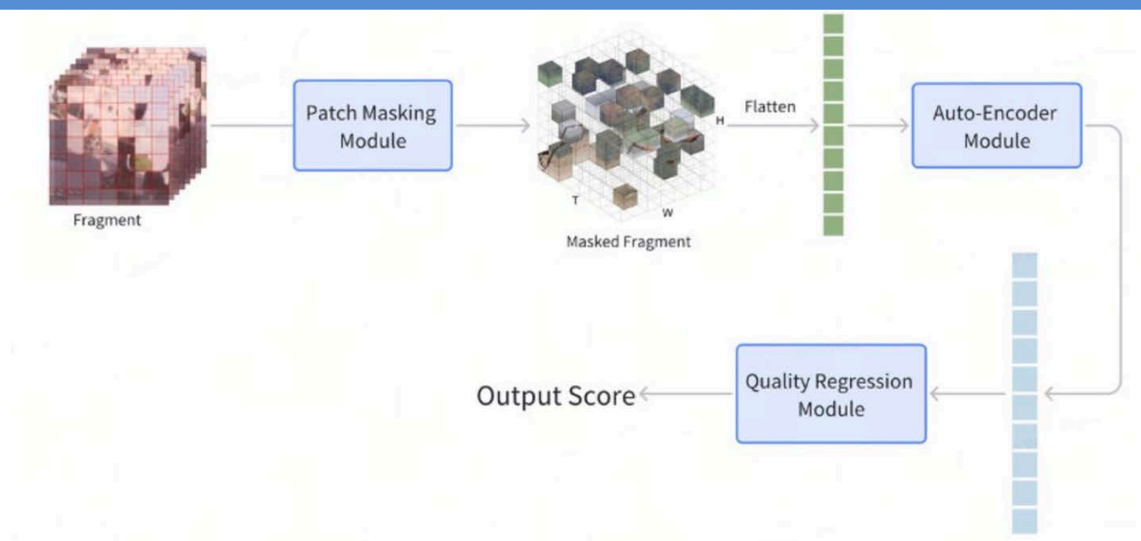
- Rapid growth in mobile and internet use increases UGC video production and diversity. The massive volume of daily video content necessitates efficient and effective quality assessment.
- Essential comprehensive quality checks throughout the video delivery process for user satisfaction.
- The high computational cost of current deep learning models, especially when applied to high frame rate and high-resolution videos, poses a significant limitation.

Datasets

- **KoNViD-1k**: A dataset with 1,200 diverse, authentic-distortion videos at 540p for real-world quality evaluation.
- **YouTube-UGC**: Contains 1,500 user-generated videos up to 2160p, representing a broad range of authentic content.
- **LSVQ**: A large-scale VQA dataset with 39,075 videos up to 1080p, including authentic and synthetic distortions.
- **LIVE-VQC**: With 585 videos up to 1080p, it provides a challenging real-world setting for quality assessment.
- **CVD2014**: A collection of 234 videos at 720p with synthetic distortions.



Methodology



- Input videos are segmented into smaller fragments by FAST-VQA for manageability.
- The video fragments are then processed into a structured format via the patch masking module.
- The auto-encoder module extracts spatiotemporal features from the unmasked regions of the fragments.
- An encoder with self-attention mechanisms processes the spatiotemporal information.
- A quality regression module synthesizes features into a score representing the video's quality.

Result

Finetune Dataset		LIVE-VQC		KoNViD-1k		CVD2014		YouTube-UGC	
Groups	Methods	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
Existing Classical	TLVQM	0.799	0.803	0.773	0.768	0.83	0.85	0.669	0.659
	VIDEVAL	0.752	0.751	0.783	0.78	NA	NA	0.779	0.773
	RAPIQUE	0.755	0.786	0.803	0.817	NA	NA	0.759	0.768
Existing Fixed Deep	VSFA	0.773	0.795	0.773	0.775	0.87	0.868	0.724	0.743
	PVQ	0.827	0.837	0.791	0.786	NA	NA	NA	NA
	GST-VQA	NA	NA	0.814	0.825	0.831	0.844	NA	NA
	ColNVQ	NA	NA	0.767	0.764	NA	NA	0.816	0.802
Ensemble C+D	CNN+TLVQM	0.825	0.834	0.816	0.818	0.863	0.88	NA	NA
	CNN+VIDEVAL	0.785	0.81	0.815	0.817	NA	NA	0.808	0.803
Full-res Swin-T features		0.799	0.808	0.841	0.838	0.868	0.87	0.798	0.796
FAST-VQA-M		0.803	0.828	0.873	0.872	0.877	0.892	0.768	0.765
FAST-VQA		0.849	0.865	0.891	0.892	**0.891**	**0.903**	**0.855**	**0.852**
DOVER		**0.86**	0.875	**0.909**	**0.906**	NA	NA	**0.89**	**0.891**
MAE-VQA（large 0.95)		**0.879**	**0.899**	0.888	0.896	0.884	0.901	0.787	0.791
MAE-VQA（huge 0.9)		**0.876**	**0.904**	**0.904**	**0.913**	**0.906**	**0.933**	**0.859**	**0.852**
MAE-VQA（huge 0.95)		0.854	**0.897**	**0.894**	**0.906**	**0.914**	**0.939**	0.832	0.822

Conclusion

- The MAE-VQA model excels in UGC video assessment, featuring innovative modules for efficient, precise quality scoring. Its use of Vision Transformers and self-attention mechanisms, combined with a unique patch masking technique, ensures reduced computational demands while enhancing accuracy.
- In future work, we will refine the MAE-VQA's masking strategy by integrating motion vectors or optical flow techniques to target perceptually significant areas, enhancing the model's focus and efficiency in predicting video quality.