

# Learning to Predict Streaming Video QoE: Distortions, Rebuffering and Memory

Christos G. Bampis, Student Member, IEEE, and Alan C. Bovik, Fellow, IEEE

SA19006037  
Jianzhao Liu

# Motivation

**For streaming applications, adaptive network strategies may involve a combination of dynamic bitrate allocation along with playback interruptions when the available bandwidth reaches a very low value.**

**Propose Video Assessment of Temporal Artifacts and Stalls (Video ATLAS): a machine learning framework where we combine a number of QoE-related features, including objective quality features, rebuffering-aware features and memory-driven features to make QoE predictions.**

# Previous Work on QoE Prediction

Impairments of Videos with Normal Playback

Due to the multiple encoding bitstream  
representations of the high-quality source content



SSIM、MS-SSIM、VMAF、STRRED

Playback interruptions

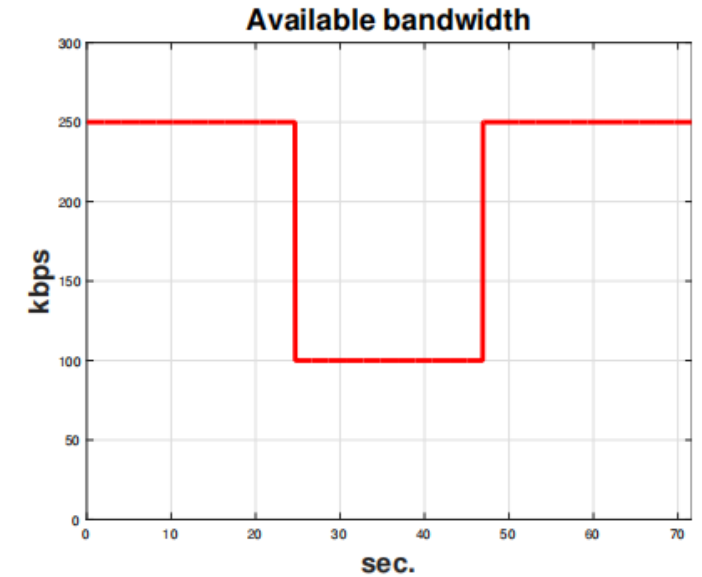
Due to throughput and buffer limitations



FTW、VsQM、SQI

# LIVE-Netflix dataset

- 8 different playout patterns (static and dynamic bitrate selection strategies together with playback interruptions) on 14 diverse video contents
- Gathered approximately 5000 subjective QoE (both continuous and retrospective) scores from 56 subjects, each participating in three 45 minute sessions.



H.264 compression



Playback interruption

# Is Objective VQA Enough ?

IQA/VQA metric	$S_q$	$S_{all}$
PSNR (IQA, FR)	0.5561	0.5152
PSNRhvs [34] (IQA, FR)	0.5841	0.5385
SSIM [11] (IQA, FR)	0.7852	0.7015
MS-SSIM [13] (IQA, FR)	0.7532	0.6800
NIQE [35] (IQA, NR)	0.3960	0.1697
VMAF [17] (VQA, FR)	0.7533	0.6097
STRRED [19] (VQA, RR)	0.7996	0.6594
GMSD [36] (IQA, FR)	0.6476	0.5812

$S_q$ : videos distorted only by video quality changes with normal playback

$S_{all}$ : all the videos in the dataset

# Learning-based Framework for QoE Prediction

## 1. Objective video quality scores (VQA)

Use pooling strategy to collapse per-frame objective quality measurements (i.e. SSIM)

## 2. Rebuffering-aware features (R1 and R2)

Use the length of each rebuffering event measured in seconds (R1) and the number of rebuffering events (R2).

## 3. Memory-related feature (M)

The time since the last rebuffering event or rate drop took place and was completed

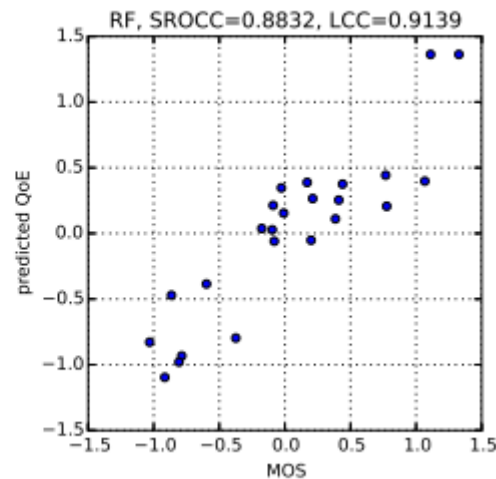
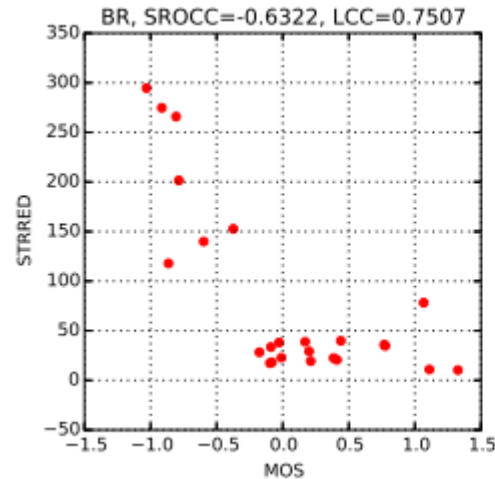
## 4. Impairment duration feature (I)

# Experiments and Results

--Regression scheme contributes to better performance (content independence )

VQA	PSNR	PSNRhvs [34]	SSIM [11]	MS-SSIM [13]	NIQE [35]	VMAF [17]	STRRED [19]	GMSD [36]	mean
BR	0.6074	0.6252	0.6748	0.6557	0.1391	0.6043	0.6348	0.6496	0.5734
Ridge	<b>0.6687</b>	<b>0.6817</b>	0.7565	0.7461	0.4130	0.6278	0.7957	<b>0.6948</b>	0.6730
Lasso	0.6496	0.6687	0.7461	0.7383	0.4191	<b>0.6409</b>	0.7983	0.6922	0.6691
SVR	0.6313	0.6417	0.8252	0.8226	0.6730	0.6026	<b>0.8704</b>	0.6878	<b>0.7193</b>
ET	0.4265	0.4387	<b>0.8547</b>	<b>0.8752</b>	<b>0.7530</b>	0.4756	0.8439	0.4527	0.6400
RF	0.4931	0.5312	0.8088	0.8154	0.6222	0.4930	0.8104	0.5417	0.6395
GB	0.4830	0.4944	0.7990	0.7899	0.5878	0.5145	0.8032	0.5000	0.6215

VQA	PSNR	PSNRhvs [34]	SSIM [11]	MS-SSIM [13]	NIQE [35]	VMAF [17]	STRRED [19]	GMSD [36]	mean
BR	0.6048	0.6534	0.7288	0.7104	0.3752	0.7561	0.7213	0.6861	0.6545
Ridge	0.8145	0.8224	0.8531	0.8517	0.5984	0.8158	0.8703	0.8254	0.8064
Lasso	<b>0.8192</b>	<b>0.8312</b>	0.8558	0.8514	0.6034	<b>0.8292</b>	0.8719	<b>0.8374</b>	0.8124
SVR	0.7939	0.8016	0.9073	0.8973	0.7633	0.7742	<b>0.9358</b>	0.8106	<b>0.8355</b>
ET	0.6325	0.6392	<b>0.9186</b>	<b>0.9289</b>	<b>0.8407</b>	0.6808	0.9088	0.6869	0.7796
RF	0.6767	0.6922	0.8905	0.8868	0.7182	0.6591	0.8770	0.7026	0.7629
GB	0.6744	0.7060	0.8661	0.8546	0.7143	0.7115	0.8678	0.7043	0.7624





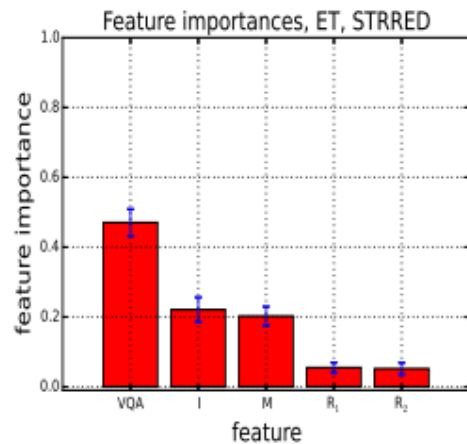
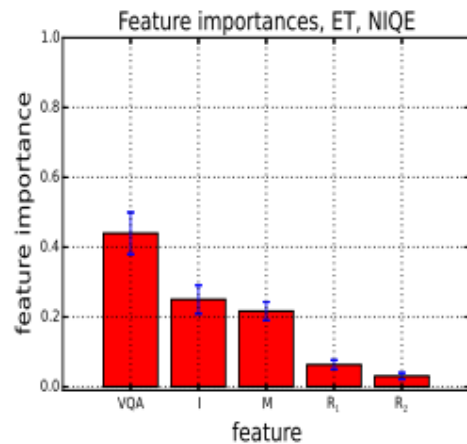
# Experiments and Results

-- Memory-related feature (M) plays an importance role (content independence )

feature subsets are indexed as follows: VQA(1), M(2), I(3),  $R_1+R_2$ (4), VQA+M(5), VQA+I(6), VQA+M+ $R_2$ (7), M+ $R_1+R_2$ (8), M+I+ $R_1+R_2$ (9), VQA+I+ $R_1+R_2$ (10), VQA+M+ $R_1+R_2$ (11) and VQA+M+I+ $R_1+R_2$ (12).

Features	1	2	3	4	5	6	7	8	9	10	11	12
Ridge	<b>0.6348</b>	0.2296	0.2700	0.3094	0.6000	0.6235	0.7870	0.4105	0.4172	0.7878	0.7735	0.7957
Lasso	<b>0.6348</b>	0.2296	0.2700	0.3243	0.6304	<b>0.6417</b>	0.7991	0.4075	0.3955	0.8013	0.7991	0.7983
SVR	0.5748	0.3807	<b>0.2758</b>	<b>0.3740</b>	0.7322	0.5878	<b>0.8183</b>	0.4210	0.4839	<b>0.8543</b>	<b>0.8122</b>	<b>0.8704</b>
ET	0.5074	0.3076	0.2345	0.2993	0.7431	0.5962	0.7496	0.3119	0.3924	0.8348	0.7574	0.8435
RF	0.5304	<b>0.3961</b>	0.2713	0.3218	<b>0.7537</b>	0.5691	0.7633	0.4126	0.4656	0.8074	0.7708	0.8096
GB	0.5691	0.3905	0.2658	0.3527	0.7461	0.6001	0.7668	<b>0.4355</b>	<b>0.4984</b>	0.8070	0.7607	0.8036

Features	1	2	3	4	5	6	7	8	9	10	11	12
Ridge	<b>0.7213</b>	0.4507	0.3049	0.2930	0.7141	0.6475	0.7610	0.4602	0.6247	0.7854	0.7590	0.8703
Lasso	<b>0.7213</b>	0.4507	0.3049	0.2956	0.7348	<b>0.6956</b>	0.7870	0.4592	0.6201	0.8055	0.7868	0.8719
SVR	0.6454	0.4325	0.3148	0.3169	<b>0.8133</b>	0.6472	<b>0.8510</b>	0.4497	0.6959	0.8945	<b>0.8392</b>	<b>0.9358</b>
ET	0.5407	0.3754	0.3110	0.3138	0.7620	0.6031	0.7596	0.3899	0.6173	<b>0.9004</b>	0.7659	0.9090
RF	0.5685	0.4451	<b>0.3528</b>	<b>0.3261</b>	0.7794	0.6024	0.7862	<b>0.4706</b>	0.6966	0.8686	0.7975	0.8742
GB	0.6287	<b>0.4514</b>	0.3514	0.3141	0.7755	0.6269	0.7904	0.4751	<b>0.7413</b>	0.8665	0.7865	0.8686





# Experiments and Results

-- content independence and pattern independence

Content independence

Method	SROCC	LCC	Best
FTW [30]	0.3403	0.2956	-
VsQM [31]	0.3120	0.2421	-
PSNR	0.6074	0.6048	-
SSIM [11]	0.6748	0.7289	-
MS-SSIM [13]	0.6557	0.7104	-
PSNR+SQI [32]	0.6565	0.6599	-
SSIM+SQI [32]	0.7565	0.8031	-
MS-SSIM+SQI [32]	0.7270	0.7731	-
PSNR+ATLAS	0.6687	0.8145	Ridge
SSIM+ATLAS	0.8547	0.9186	ET
MS-SSIM+ATLAS	<b>0.8752</b>	<b>0.9289</b>	ET

Pattern independence

Method	SROCC	LCC	Best
PSNR	0.4945	0.5312	-
PSNR+SQI [32]	0.4989	0.5340	-
PSNR+ATLAS	0.4945	0.5321	Ridge
SSIM [11]	0.6615	0.7947	-
SSIM+SQI [32]	0.6791	0.7927	-
SSIM+ATLAS	0.7143	0.8650	RF
MS-SSIM [13]	0.6659	0.7982	-
MS-SSIM+SQI [32]	0.6835	0.7955	-
MS-SSIM+ATLAS	0.6961	0.8345	GB
NIQE	0.4681	0.4107	-
NIQE+ATLAS	0.6447	0.6541	RF
VMAF	0.3890	0.4486	-
VMAF+ATLAS	0.7415	0.7075	RF
STRRED	0.8066	0.7848	-
STRRED+ATLAS	<b>0.8198</b>	<b>0.7923</b>	Ridge
GMSD	0.4989	0.5545	-
GMSD+ATLAS	0.5256	0.6679	RF

# Experiments and Results

-- generalizability

Method	SROCC	LCC	Best
FTW [30]	0.3290	0.3358	-
VsQM [31]	0.2358	0.3324	-
PSNR	0.6894	0.6875	-
SSIM [11]	0.8172	0.8544	-
MS-SSIM [13]	0.7986	0.8345	-
SSIMplus [49]	0.8025	0.8414	-
PSNR+SQI [32]	0.7800	0.7535	-
SSIM+SQI [32]	0.9085	0.9028	-
MS-SSIM+SQI [32]	0.8891	0.8808	-
SSIMplus+SQI [32]	0.9103	0.9012	-
PSNR+ATLAS	0.7799	0.7510	SVR
SSIM+ATLAS	<b>0.9142</b>	<b>0.9097</b>	SVR
MS-SSIM+ATLAS	0.8955	0.8880	Lasso
SSIMplus+ATLAS	0.9084	0.8981	Ridge

Training and Testing for Waterloo

Method	SROCC	LCC	Best
FTW [30]	0.3352	0.2900	-
VsQM [31]	0.3236	0.2374	-
PSNR	0.5152	0.5073	-
SSIM [11]	0.7015	0.7219	-
MS-SSIM [13]	0.6800	0.7104	-
PSNR+SQI [32]	0.5904	0.5905	-
SSIM+SQI [32]	0.7451	0.7070	-
MS-SSIM+SQI [32]	0.7239	0.6848	-
PSNR+ATLAS	0.6155	0.6116	SVR
SSIM+ATLAS	<b>0.8203</b>	<b>0.7813</b>	Lasso
MS-SSIM+ATLAS	0.8000	0.7670	Lasso

training on Waterloo and testing on LIVE-Netflflix

Method	SROCC	LCC	Best
FTW [30]	0.3154	0.3313	-
VsQM [31]	0.2259	0.3233	-
PSNR	0.6715	0.6587	-
SSIM [11]	0.8177	0.8408	-
MS-SSIM [13]	0.7928	0.8168	-
PSNR+SQI [32]	0.7492	0.7316	-
SSIM+SQI [32]	0.9009	0.8897	-
MS-SSIM+SQI [32]	0.8807	0.8652	-
PSNR+ATLAS	0.7439	0.7254	SVR
SSIM+ATLAS	<b>0.9090</b>	<b>0.8963</b>	Lasso
MS-SSIM+ATLAS	0.8888	0.8716	Lasso

training on LIVE-Netflflix and testing on Waterloo

# Conclusion

- The recency/memory effects contributes to retrospective QoE evaluation
- No general model for different datasets (only a framework)
- Design continuous time QoE models in the future

Thanks