MVAD: A Multiple Visual Artifact Detector for Video Streaming Supplementary Material

Chen Feng†, Duolikun Danier†, Fan Zhang†, Alex Mackin‡, Andrew Collins‡, David Bull†
† Visual Information Lab, University of Bristol, BS1 5DD, United Kingdom
‡Amazon Prime Video, 1 Principal Place, Worship Street, London, EC2A 2FA, United Kingdom
{chen.feng, duolikun.danier, fan.zhang, dave.bull}@bristol.ac.uk,
{acmackin, accllin}@amazon.co.uk

A. ROC Plots

Figure 1 and Figure 3 show the ROC plots for eight artifact detection methods on each artifact tested based on the Maxwell and BVI-Artifact database. The ROC (Receiver Operating Characteristic) curves provide a visual representation of the performance of these methods, illustrating the trade-off between the true positive rate (TPR) and the false positive rate (FPR) across different threshold settings.

B. Visualization of the Guided Mask in ADFE

Figure 2 showcases the influence of the guided mask $\mathbf{M}$ in the ADFE module, which does emphasize the regions with visual artifacts.

C. Artifacts Synthesis Methods

Table 1 provides a summary of the synthesis methods used to generate various video artifacts. In the Augmented Database, different parameters were employed to produce artifacts with four visibility levels.

D. Broader Impacts

The development and implementation of the MVAD framework for detecting multiple visual artifacts in streamed videos have the following impacts.

On the positive side, the MVAD system supports automatic inspection of video content to identify visual artifacts, significantly reducing labor costs and increasing the efficiency of quality control processes in video streaming services. MVAD can also potential improve the quality of streamed videos by incorporating with video enhancement methods, leading to a better viewing experience for end users.

The implementation of MVAD can also result in negative impacts. The training and deployment of sophisticated machine learning models such as MVAD require substantial computational resources, which can lead to increased energy consumption and negative environmental impact. Further model complexity reduction can alleviate this issue, and it remains our future work.

E. License of Code and Data

Table 2 summarizes the license associated with the code and data used and generated in this work.

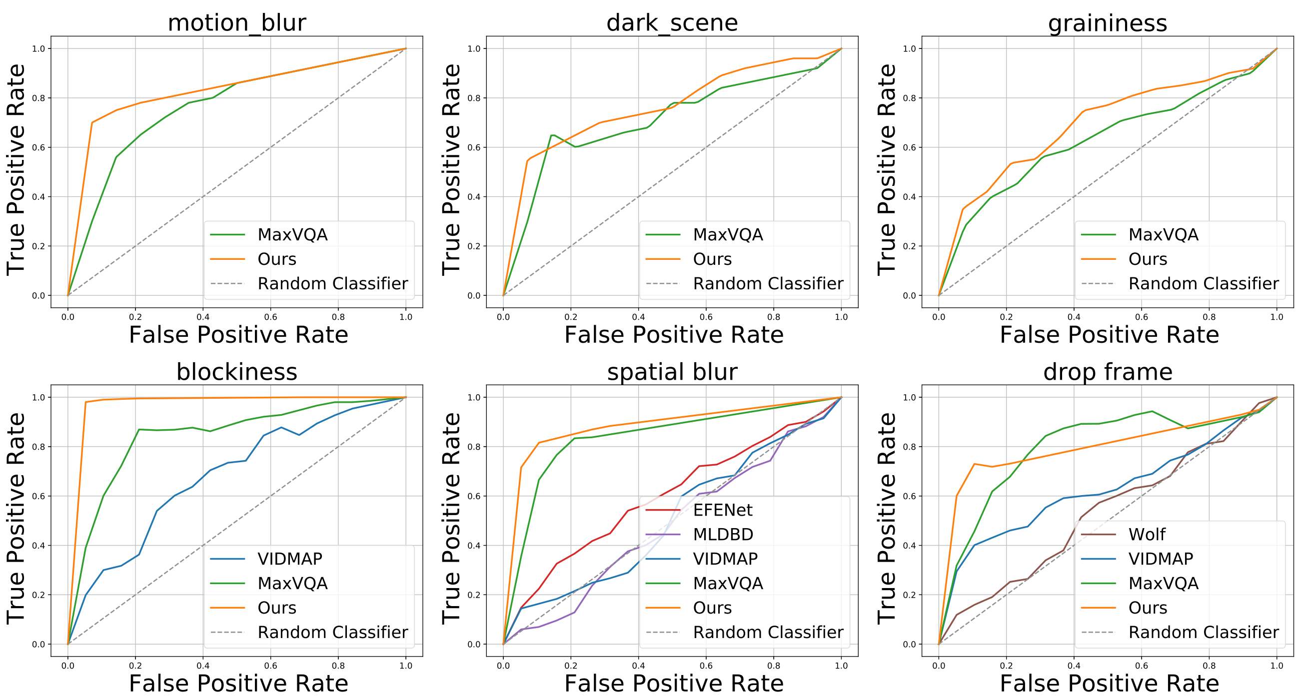

Figure 1. The ROC curves for different artifact categories in the Maxwell [13] database.

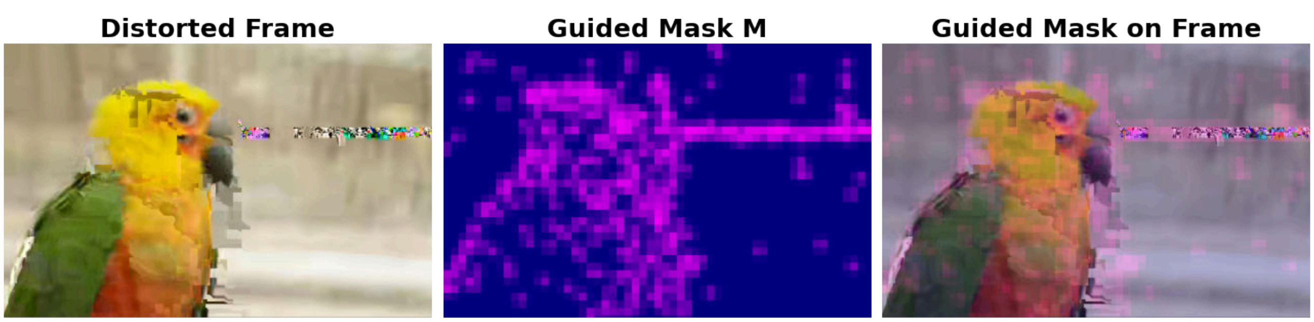

Figure 2. Visualization of the guided mask $\mathbf{M}$ generated by the ADFE module.

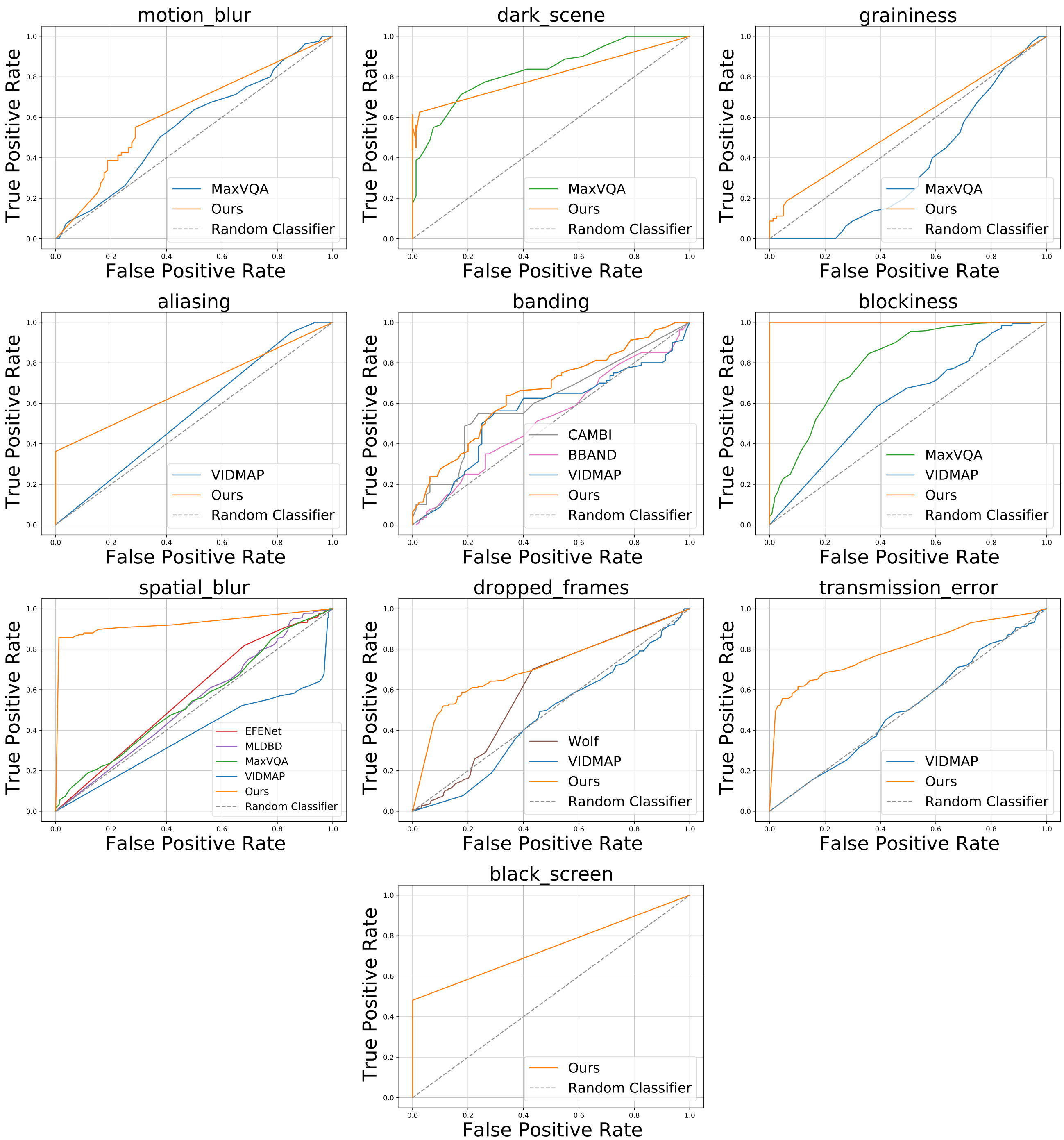

Figure 3. The ROC curves for different artifact categories in the BVI-Artifact [1] database.

Artifacts	Synthesis Methods	Levels	Parameters
Aliasing	Spatial re-sampling [2]	Very noticeable	Sampling ratio=4
		Noticeable	Sampling ratio=3
		Subtle	Sampling ratio=2
		Very subtle	Sampling ratio=1.5
Banding	Quantization [2]	Very noticeable	Quantization ratio=5
		Noticeable	Quantization ratio=4
		Subtle	Quantization ratio=3
		Very subtle	Quantization ratio=2
Dark Scene	Brightness/contrast adjustment [3]	Very noticeable	Decrease ratio=4
		Noticeable	Decrease ratio=3
		Subtle	Decrease ratio=2
		Very subtle	Decrease ratio=1.5
Motion Blur	Convolution (motion blur) [8]	Very noticeable	Frame numbers=16
		Noticeable	Frame numbers=12
		Subtle	Frame numbers=8
		Very subtle	Frame numbers=4
Graininess	Gaussian noise [2]	Very noticeable	standard deviation=50
		Noticeable	standard deviation=25
		Subtle	standard deviation=10
		Very subtle	standard deviation=5
Blockiness	JPEG compression [2]	Very noticeable	Quantization parameter= 47
		Noticeable	Quantization parameter= 42
		Subtle	Quantization parameter= 37
		Very subtle	Quantization parameter= 32
Frame Drop	Random frame dropping [12]	Very noticeable	Frame length = 16
		Noticeable	Frame length = 12
		Subtle	Frame length = 8
		Very subtle	Frame length = 4
Spatial Blur	Convolution (Gaussian blur) [2]	Very noticeable	Kernel size=9
		Noticeable	Kernel size=7
		Subtle	Kernel size=5
		Very subtle	Kernel size=3
Transmission Error	Packet loss and error concealment [2]	Very noticeable	bitstream filter = 4M
		Noticeable	bitstream filter = 2M
		Subtle	bitstream filter = 1M
		Very subtle	bitstream filter = 0.5M
Black Screen	Random black frames replacement	Very noticeable	Frame length = 16
		Noticeable	Frame length = 12
		Subtle	Frame length = 8
		Very subtle	Frame length = 4

Table 1. Artifacts synthesis methods used for multiple artifacts generation.

Code/Data	Size	Dataset URL	License/Terms of Use
		Training Datasets	
BVI-DVC [6]	800	https://fan-aaron-zhang.github.io/BVI-DVC/	Academic research.
BVI-CC [4]	90	https://fan-aaron-zhang.github.io/BVI-CC/	Academic research.
NFLX-public [5]	70	https://github.com/Netflix/vmaf/blob/master/resource/doc/datasets.md	Academic research.
LIVE-HFR [7]	88	https://fan-aaron-zhang.github.io/BVI-HFR/	Academic research.
Adobe240 [9]	133	https://www.cs.ubc.ca/labs/imager/tr/2017/DeepVideoDeblurring/	MIT license.
		Test Datasets	
BVI-Artifact [1]	480	https://chenfeng-bristol.github.io/BVI-Artefact/	Academic research.
Maxwell [13]	4,543	https://github.com/VQAssessment/ExplainableVQA	MIT license.
		Code of Benchmark Methods	
MaxVQA [13]	-	https://github.com/VQAssessment/ExplainableVQA	MIT license.
VIDMAP [2]	-	https://github.com/utlive/VIDMAP	Academic research.
CAMBI [10]	-	https://github.com/Netflix/vmaf/blob/master/resource/doc/cambi.md	BSD+Patent.
BBAND [11]	-	https://github.com/google/bband-adaband	Apache-2.0 license.
EFENet [14]	-	https://github.com/wdzhao123/DENets	No licence.
MLDBD [15]	-	https://github.com/wdzhao123/MLDBD	No licence.
Wolf *el at.* [12]	-	https://its.ntia.gov/umbraco/surface/download/publication?reportNumber=TR-09-456.pdf	Academic research.
		Code and Data generated in this work	
MVAD	-	https://chenfeng-bristol.github.io/MVAD/	CC-BY-4.0
Training data [6]	50,800	https://chenfeng-bristol.github.io/MVAD/	Academic research.

Table 2. License information for the Code and datasets used and generated in this work.

References

[1] Chen Feng, Duolikun Danier, Fan Zhang, and David Bull. BVI-Artefact: An artefact detection benchmark dataset for streamed videos. *arXiv preprint arXiv:2312.08859*, 2023. 3, 5

[2] Todd R Goodall and Alan C Bovik. Detecting and mapping video impairments. *IEEE Trans. on Image Processing*, 28(6):2680–2691, 2018. 4, 5

[3] Xiaojie Guo, Yu Li, and Haibin Ling. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 4

[4] Angeliki Katsenou, Fan Zhang, Mariana Afonso, Goce Dimitrov, and David R Bull. BVI-CC: A dataset for research on video compression and quality assessment. *Frontiers in Signal Processing*, 2:874200, 2022. 5

[5] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara. Toward a practical perceptual video quality metric. *The Netflix Tech Blog*, 2016. 5

[6] Di Ma, Fan Zhang, and David R Bull. BVI-DVC: A training database for deep video compression. *IEEE Transactions on Multimedia*, 24:3847–3858, 2021. 5

[7] Pavan C Madhusudana, Xiangxu Yu, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C Bovik. Subjective and objective quality assessment of high frame rate videos. *IEEE Access*, 9:108069–108082, 2021. 5

[8] Wang Shen, Wenbo Bao, Guangtao Zhai, Li Chen, Xiongkuo Min, and Zhiyong Gao. Blurry video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5114–5123, 2020. 4

[9] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1279–1288, 2017. 5

[10] Pulkit Tandon, Mariana Afonso, Joel Sole, and Lukáš Krasula. CAMBI: Contrast-aware multiscale banding index. In *2021 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2021. 5

[11] Zhengzhong Tu, Jessie Lin, Yilin Wang, Balu Adsumilli, and Alan C Bovik. BBAND index: A no-reference banding artifact predictor. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2712–2716. IEEE, 2020. 5

[12] Stephen Wolf. A no reference (NR) and reduced reference (RR) metric for detecting dropped video frames. Technical report, Institute for Telecommunication Sciences, 2008. 4, 5

[13] Haoning Wu, Erli Zhang, Liang Liao, Chaofeng Chen, Jingwen Hou, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. Towards explainable in-the-wild video quality assessment: A database and a language-prompted approach. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, page 1045–1054. Association for Computing Machinery, 2023. 2, 5

[14] Wenda Zhao, Xueqing Hou, You He, and Huchuan Lu. Defocus blur detection via boosting diversity of deep ensemble networks. *IEEE Trans. on Image Processing*, 30:5426–5438, 2021. 5

[15] Wenda Zhao, Fei Wei, Haipeng Wang, You He, and Huchuan Lu. Full-scene defocus blur detection with DeFBD+ via multi-level distillation learning. *IEEE Trans. on Multimedia*, 25:9228–9240, 2023. 5