SALIENCY-AWARE END-TO-END LEARNED VARIABLE-RATE 360-DEGREE IMAGE COMPRESSION SUPPLEMENTARY MATERIAL

Oğuzhan Güngördü, A. Murat Tekalp

Department of Electrical and Electronics Engineering, Koc University, 34450 Istanbul, Turkey

1. OVERVIEW

In this supplementary material, we provide additional saliency results and visual quality comparisons of the decoded images that complement our main paper. Firstly, we extend the ablation studies introduced in Sec. 2, exploring the effects of various baseline configurations for ODI compression. Furthermore, we include comprehensive benchmark results from the Salient 360! Grand Challenge ICME2017 [1] in Sec. 3. Then, we extend our visual saliency comparison in Sec. 3 by incorporating models such as BMS [2], SALICON [3], BMS360 [4], and GVBS360 [4]. Furthermore, due to the computational constraints associated with processing 360° images at original resolution using LIC360 [5], we present a detailed comparison of our model with LIC360 in Sec. 4.

2. ABLATION STUDIES

2.1. Selection of Baseline Compression Architecture

In our evaluation of baseline effects for 360° image compression, the ELIC and Cheng20 architectures were trained with latent masking and the Sal-MSE loss, as detailed in the last two rows of Tab. 1. It shows ELIC's inherent strength, outperforming Cheng20 with the same enhancements. However, the first and last rows of Tab. 1 reveal that Cheng20 exceeds unmodified ELIC in the SAL-PSNR metric in our proposed architecture. It highlights our model's adaptability; despite ELIC's 2D compression superiority, our enhancements enable Cheng20 to outperform ELIC in the salient regions of 360° images, underscoring our framework's potential in enhancing 2D codecs for 360° images.

Method	WS-PSNR		SAL-PSNR	
	BD-psnr	BD-rate	BD-psnr	BD-rate
Base 1	0 dB	0	0 dB	0
Base 1+Sal-MSE	0.20 dB	-5.30%	0.26 dB	-6.30%
Base 1+Masking	0.18 dB	-4.77%	0.29 dB	-7.11%
Base 1+Sal-MSE+Masking	**0.42 dB**	**-10.21%**	**0.57 dB**	**-12.85%**
Base 2+Sal-MSE+Masking	-0.14 dB	4.07%	0.12 dB	-3.24%

Table 1: Compression ablation study on Salient360! 2017 test set using ELIC [6] (Base 1) and Cheng20 [7] (Base 2) as baselines.

3. COMPARISON WITH THE SOTA: SALIENCY DETECTION

Our proposed saliency architecture's performance is visually benchmarked against eight models in Fig. 1. This set includes two leading 2D models, BMS [2] and SALICON [3], along with six 360° models: BMS360 [4], GVBS360 [4], SalNet360 [8], MV-SalGAN360 [9], ATSal [10], and MRGAN360 [11]. These models were selected based on the availability of their source code. We also include extended benchmark results from the Salient 360! Grand Challenge ICME2017 [1] in Tab. 2. Overall, our model excels in the CC metric, while MRGAN360 stands out in NSS and AUC-J, and MV-SalGAN360 in KLD. However, our model remains competitive across all metrics.

Method	KLD ↓	CC ↑	NSS ↑	AUC-J ↑
Maugey *et al.* [13]	0.585	0.448	0.506	0.644
Zhang *et al.* [14]	—	0.409	0.699	0.659
SalNet360 [8]	0.458	0.548	0.755	0.701
SalGAN [15]	1.236	0.452	0.810	0.708
Startsev *et al.* [16]	0.42	0.62	0.81	0.72
GBVS360 [4]	0.698	0.527	0.851	0.714
BMS360 [4]	0.599	0.554	0.936	0.736
SalGAN&FSM [17]	0.896	0.512	0.910	0.723
Zhu *et al.* [18]	0.481	0.532	0.918	0.734
Ling *et al.* [19]	0.477	0.550	0.939	0.736
SalGAN360 [20]	0.431	0.659	0.971	0.746
MV-SalGAN360 [9]	**0.363**	0.662	0.978	0.747
MRGAN360 [11]	0.401	0.658	**1.09**	**0.784**
SalBiNet360 [21]	0.402	0.661	0.975	0.746
ATSal [10]	0.449	0.630	0.865	0.693
Ours	0.406	**0.669**	0.981	0.737

Table 2: Extended comparison of saliency detection methods on the test set of Salient360! 2017 Benchmark [1].

In extended visual comparison in Fig. 1, we included additional results from BMS [2], SALICON [3], BMS360 [4], and GVBS360 [4]. These added results demonstrate that these models perform poorly in 360° saliency detection. Notably, BMS360 performs similarly to BMS, with the key difference being that it does not overestimate saliency maps as much. This observation underscores the necessity for specific 360° saliency models instead of merely extending 2D models for 360° applications.

4. VISUAL QUALITY

In this section, we present a visual comparison of our compression model with LIC360 [5], focusing on the visual quality of the decoded images. Due to the high computational cost associated with processing 360° images at their original resolution, both our model and LIC360 [5] compress images that are center cropped to 3072×1536 before zero padding to manage memory constraints. This preprocessing approach ensures a fair comparison between the two models.

As demonstrated in Fig. 2, our method shows superior performance in preserving high-frequency details within salient regions

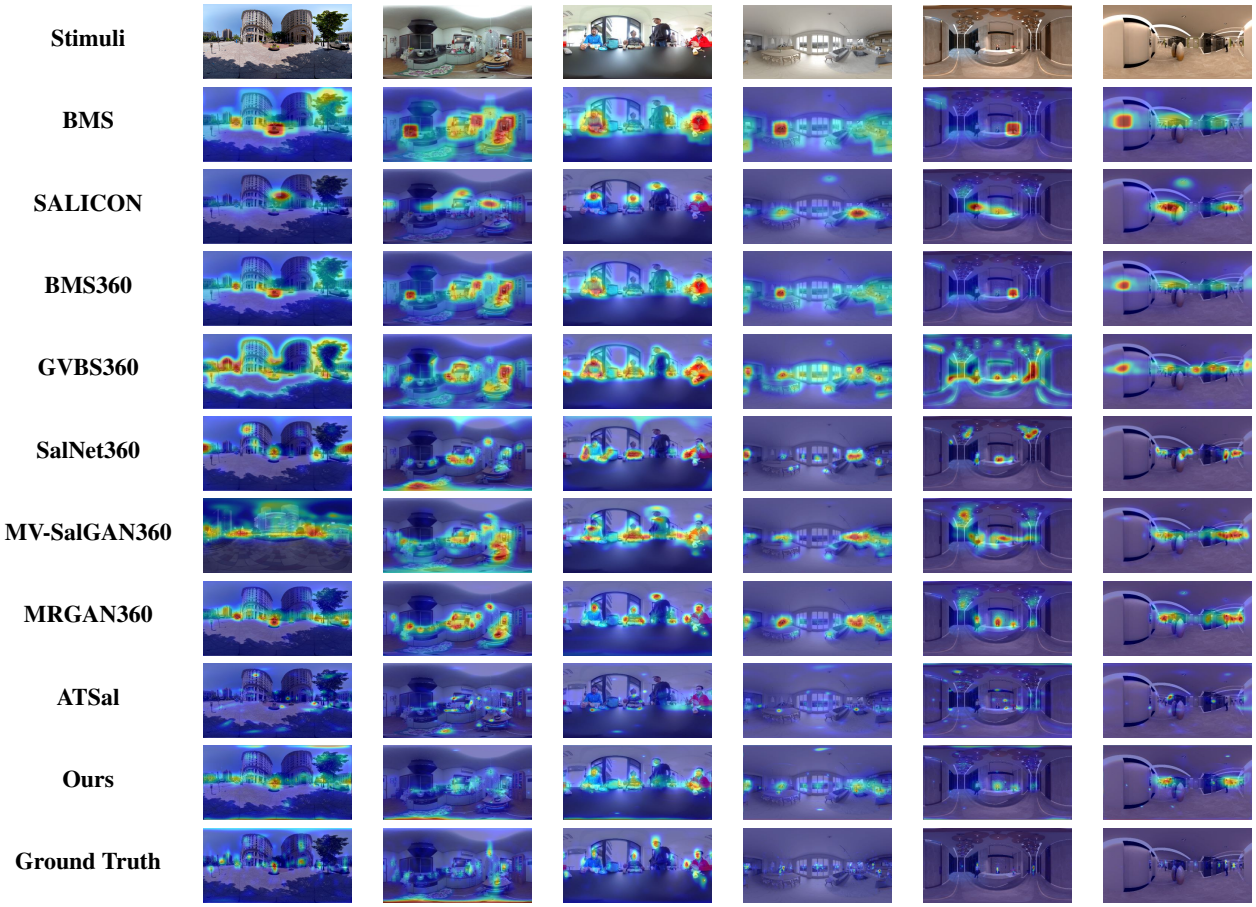


Fig. 1: Extended visual comparison with other 2D and 360° saliency models: the first three columns represent the Salient360! 2017 test set [1], and the last three columns the Saliency in VR test set [12].

(as denoted by the green boxes), such as in clothing, faces, and salient background elements, when compared to LIC360 [5]. This is evident through less blur and more refined textures in the crops of the decompressed image. Moreover, our model maintains competitive performance in nonsalient regions (indicated by the red box), highlighting that our approach, while significantly enhancing visual quality in areas of interest, also preserves overall image integrity and does not sacrifice quality in less critical areas as much. This balance underscores our model's ability to deliver top performance in salient regions without sacrificing quality in nonsalient regions.

5. REFERENCES

[1] University of Nantes, "Salient360!: Visual attention modeling for 360 images grand challenge," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2017, pp. 35–42.

[2] J. Zhang and S. Sclaroff, "Exploiting surroundedness for saliency detection: A boolean map approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 889–902, may 2016.

[3] M. Jiang, S. Huang, J. Duan, and Q. Zhao, "Salicon: Saliency in context," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[4] P. Lebreton and A. Raake, "Gbvs360, bms360, prosal: Extending existing saliency prediction models from 2d to omnidirectional images," *Signal Processing: Image Communication*, vol. 69, pp. 69–78, 2018.

[5] M. Li, J. Li, S. Gu, F. Wu, and D. Zhang, "End-to-end optimized 360° image compression," *IEEE Trans. on Image Processing*, vol. 31, pp. 6267–6281, 2022.

[6] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, "Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 5718–5727.

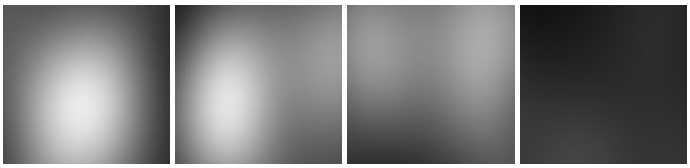
[7] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learned image compression with discretized gaussian mixture likelihoods and attention modules," in *2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7936–7945.

[8] R. Monroy, S. Lutz, T. Chalasani, and A. Smolic, "Salnet360: Saliency maps for omni-directional images with cnn," *Signal Processing: Image Communication*, vol. 69, pp. 26–34, 2018.

[9] F. Y. Chao, L. Zhang, W. Hamidouche, and O. Déforges, "A multi-fov viewport-based visual saliency model using adaptive weighting losses for 360° images," *IEEE Trans. on Multimedia*, vol. 23, pp. 1811–1826, 2021.

Original image (P26)

Original crops



Saliency maps

Ours (0.129/29.97/0.833)

LIC360 (0.130/27.92/0.791)

Fig. 2: Visual comparison of four crops from "P26" of Salient 360! 2017 [1] test set decoded by ours and LIC360 [5] codecs around 0.13 bpp. In the original image (at the top), green boxes denote salient regions and a red box denotes a nonsalient region. The crops illustrate these regions as highlighted by the boxes. The metrics under subfigures are (bpp↓ /SAL-PNSR↑ /WS-SSIM↑).

[10] Y. Dahou, M. Tliba, K. McGuinness, and N. O'Connor, "Atsal: An attention based architecture for saliency prediction in 360 videos," in *Int. Conf. on Pattern Recognition (ICPR)*. Springer, 2021, pp. 305–320.

[11] P. Gao, X. Chen, R. Quan, and W. Xiang, "Mrgan360: Multi-stage recurrent generative adversarial network for 360 degree image saliency prediction," *ArXiv*, vol. abs/2303.08525, 2023.

[12] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, "How do people explore virtual environments?," *IEEE Trans. on Visualization and Computer Graphics*, 2017.

[13] T. Maugey, O. Le Meur, and Z. Liu, "Saliency-based navigation in omnidirectional image," in *2017 IEEE 19th Int. Workshop on Multimedia Signal Processing (MMSP)*, 2017, pp. 1–6.

[14] Z. Zhang, Y. Xu, J. Yu, and S. Gao, "Saliency detection in 360° videos," in *The European Conf. on Computer Vision (ECCV)*, September 2018.

[15] J. Pan, C. Canton-Ferrer, K. McGuinness, N. O'Connor, J. Torres, E. Sayrol, and X. Giró i Nieto, "Salgan: Visual saliency prediction with generative adversarial networks," in *CVPR 2017 Scene Understanding Workshop (SUNw)*, Honolulu, Hawaii, USA, 2017.

[16] M. Startsev and M. Dorr, "360-aware saliency estimation with conventional image saliency predictors," *Signal Processing: Image Communication*, vol. 69, pp. 43–52, 2018.

[17] A. De Abreu, C. Ozcinar, and A. Smolic, "Look around you: Saliency maps for omnidirectional images in vr applications," in *2017 Ninth Int. Conf. on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1–6.

[18] Y. Zhu, G. Zhai, X. Min, and J. Zhou, "The prediction of saliency map for head and eye movements in 360 degree images," *IEEE Trans. on Multimedia*, vol. 22, no. 9, pp. 2331–2344, 2020.

[19] J. Ling, K. Zhang, Y. Zhang, D. Yang, and Z. Chen, "A saliency

prediction model on 360 degree images using color dictionary based sparse representation," *Signal Processing: Image Communication*, vol. 69, pp. 60–68, 2018.

[20] F. Y. Chao, L. Zhang, W. Hamidouche, and O. Deforges, "Salgan360: Visual saliency prediction on 360 degree images with generative adversarial networks," in *2018 IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW)*, 2018, pp. 01–04.

[21] D. Chen, C. Qing, X. Xu, and H. Zhu, "Salbinet360: Saliency prediction on 360° images with local-global bifurcated deep network," in *2020 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 92–100.