ST360IQ: NO-REFERENCE OMNIDIRECTIONAL IMAGE QUALITY ASSESSMENT WITH SPHERICAL VISION TRANSFORMERS

Nafiseh Jabbari Tofighi⋆, Mohamed Hedi Elfkir†, Nevrez Imamoglu∗, Cagri Ozcinar‡, Erkut Erdem†, Aykut Erdem⋆

⋆ Koc University KUIS AI Center, † Hacettepe University, ∗ AIST Japan, ‡ MSK.AI

Introduction

Omnidirectional images

- Omnidirectional images as a new form of visual data
- Omnidirectional images allow for new applications for various fields such as virtual reality, robotics, and surveillance
- Need for reliable quality assessment metrics for omnidirectional images

Challenges in omnidirectional image quality assessment

- Unique characteristics of omnidirectional images such as non-uniform resolution, distortion, and non-linearity
- Lack of widely accepted quality assessment metrics for omnidirectional images

Related Works

- CNN-based 360 image quality assessment:
  - MC360IQA[1]
  - VGCN[2]

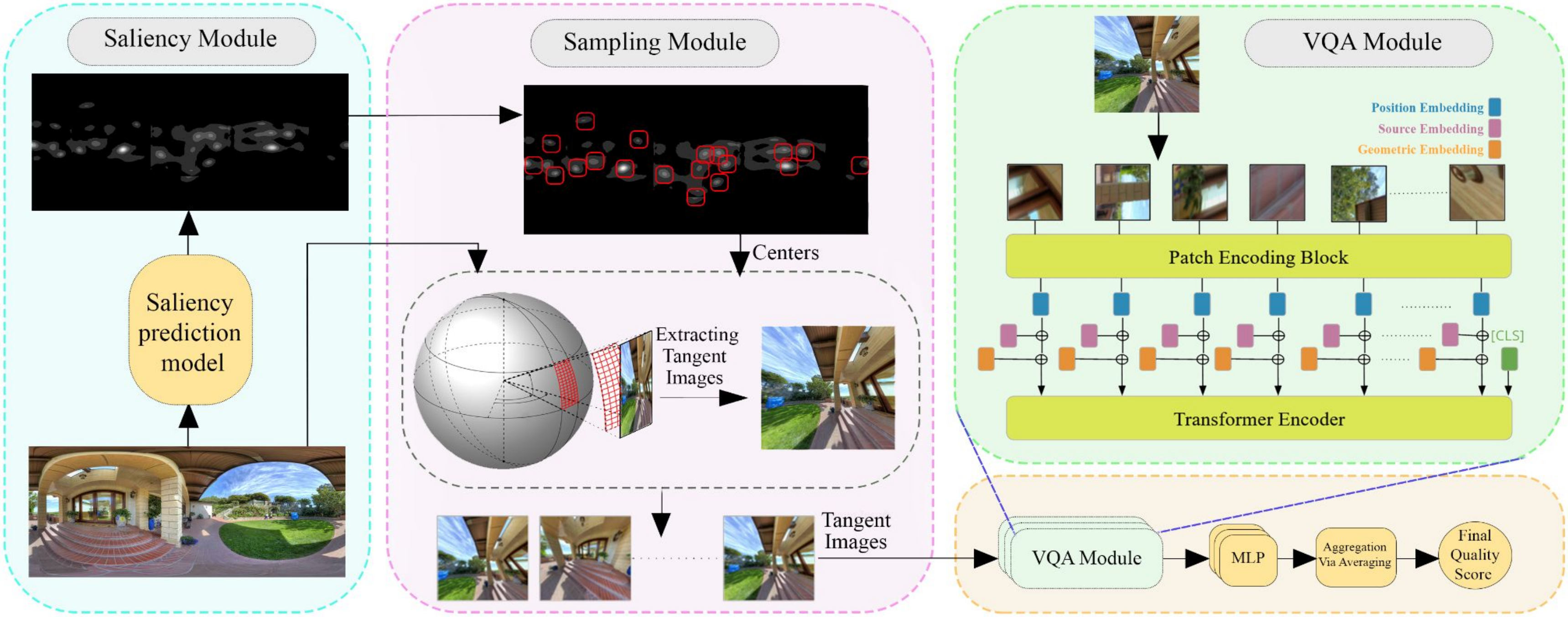
- Vision Transformer based quality assessment for natural images:
  - MuSIQ[3]

1. Wei Sun, Xiongkuo Min, Guangtao Zhai, Ke Gu, Huiyu Duan, and Siwei Ma, “Mc360iqa: A multi-channel cnn for blind 360-degree image quality assessment,” IEEE J. Sel. Top. Signal Process., vol. 14, no. 1, pp. 64–77, 2020.
2. Jiahua Xu, Wei Zhou, and Zhibo Chen, “Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks,” IEEE Trans. Circuits Syst. Video Technol., 2020
3. Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang, “MUSIQ: Multi-scale image quality transformer,” in ICCV, 2021, pp. 5148–5157.

Motivation

- Utilizing the spherical vision transformer for evaluating quality of omnidirectional images
- Combining both groups of existing methods, 360IQA and ViT
- Enjoying saliency information to increase efficiency of vision transformers as computationally expensive models

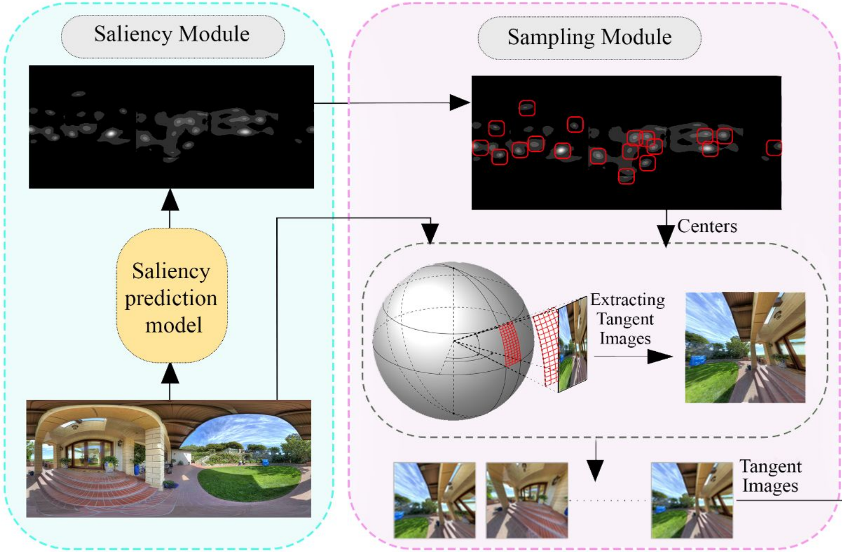
Model Overview

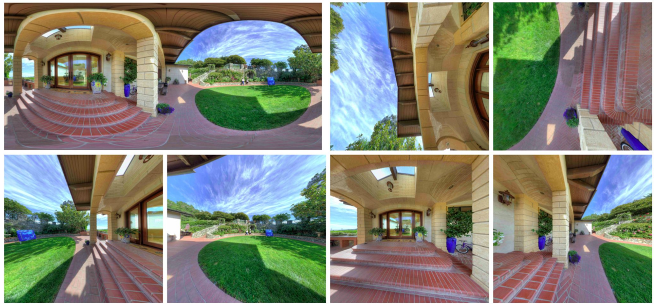


Methodology (1/3)

Sampling Module:

1. Saliency Module
2. Salient Centers
3. Tangent Images

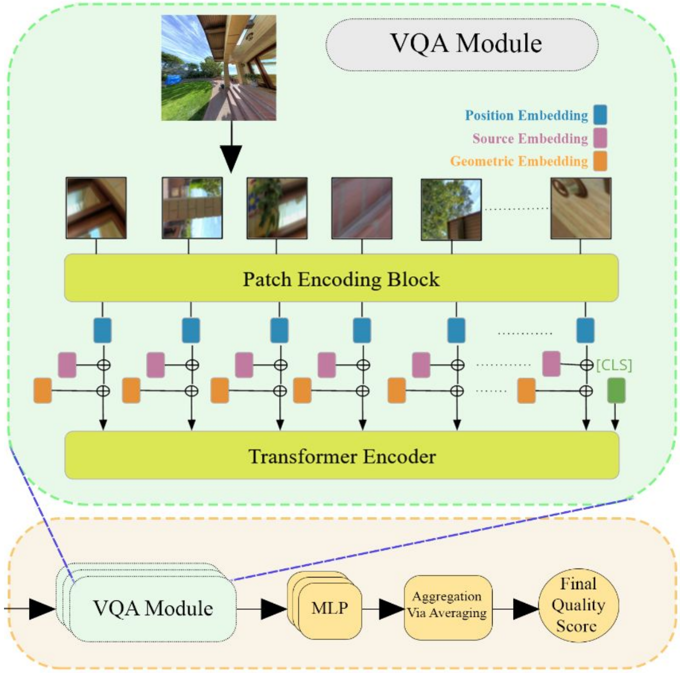




Methodology (2/3)

Patch Encoder:

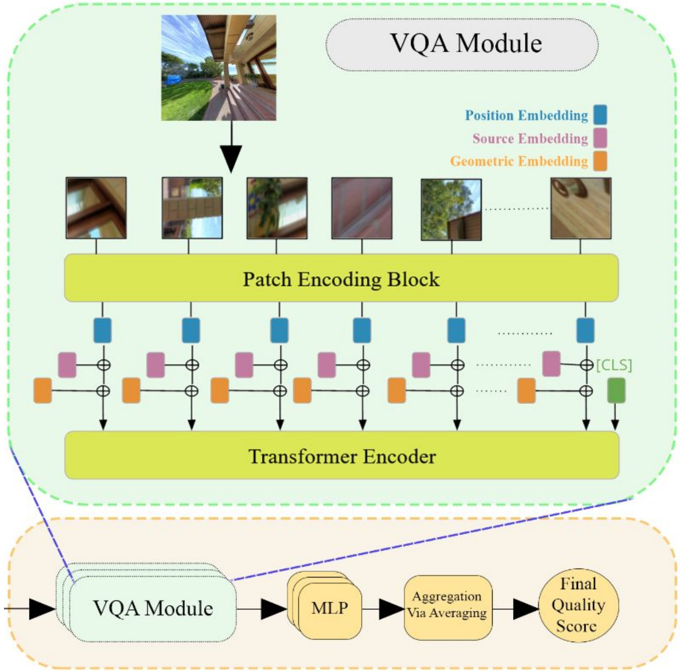
1. ResNet-50
2. Model Embedding
3. Final Prediction



Methodology (3/3)

Model Embeddings:

- Position Embedding
- Geometric Embedding
- Source Embedding



Datasets

- **CVIQ:**
  - 528 compressed images
  - 16 reference images
  - compressed by JPEG,H.264/AVC, and H.265/HEVC

- **OIQA:**
  - 16 reference images, and 320 distorted images
  - 4 distortion types(Gaussian blur, Gaussian noise,
    JPEG compression, and JPEG2000 compression)

Evaluation Criteria

- Spearman’s Rank Order Correlation Coefficient(SROCC),
- Pearson’s Linear Correlation Coefficient (PLCC)
- Root Mean Squared Error (RMSE)

Performance Comparison(CVIQ)

Table 1. Quantitative comparison of ST360IQ against the state-of-the-art on CVIQ. Bold scores indicate the best performances.

		JPEG			H.264/AVC			H.265/HEVC			Overall		
	Method	PLCC↑	SRCC↑	RMSE↓	PLCC↑	SRCC↑	RMSE↓	PLCC↑	SRCC↑	RMSE↓	PLCC↑	SRCC↑	RMSE↓
Full Reference	PSNR	0.75	0.76	10.66	0.66	0.66	10.06	0.60	0.57	9.47	0.65	0.68	10.65
	SSIM [16]	0.98	0.95	3.41	0.88	0.86	6.39	0.85	0.82	6.28	0.90	0.87	5.95
	FSIM [17]	0.98	0.96	3.37	0.95	0.94	4.34	0.95	0.95	3.67	0.95	0.94	4.50
	MS_SSIM [18]	0.95	0.89	5.06	0.75	0.73	8.78	0.73	0.72	8.02	0.83	0.78	7.88
	IW_SSIM [19]	0.98	0.96	3.03	0.94	0.94	4.37	0.95	0.95	3.63	0.91	0.90	5.71
	SR-SIM [20]	0.97	0.94	3.92	0.89	0.86	6.19	0.91	0.89	4.99	0.88	0.86	6.52
	GMSD [21]	0.96	0.91	4.28	0.73	0.72	9.07	0.81	0.81	6.96	0.82	0.79	8.03
	VSI [13]	0.96	0.91	4.59	0.87	0.85	6.67	0.86	0.84	5.97	0.89	0.85	6.41
	HaarPSI [22]	0.97	0.95	3.63	0.87	0.85	6.55	0.89	0.88	5.27	0.90	0.87	5.98
	LPIPS [3]	0.93	0.85	6.07	0.96	0.96	3.77	0.95	0.95	3.85	0.92	0.91	5.53
	DISTS [1]	0.96	0.91	4.74	0.97	0.97	3.34	0.96	0.96	3.34	0.94	0.93	4.90
	MDSI [23]	0.98	0.95	3.41	0.91	0.89	5.44	0.93	0.92	4.30	0.92	0.90	5.46
No Ref.	BRISQUE [24]	0.86	0.83	8.31	0.81	0.90	13.37	0.62	0.79	9.27	0.75	0.78	9.21
	MUSIQ [6]	0.94	0.84	5.55	0.90	0.84	5.85	0.85	0.81	6.24	0.89	0.81	6.43
	MC360IQA [8]	0.96	0.96	4.30	0.96	0.96	3.65	0.90	0.91	5.00	0.95	0.95	4.65
	VGCN [9]	**0.99**	**0.98**	**2.50**	0.97	0.97	3.15	0.94	0.95	3.99	0.96	0.96	3.67
	ST360IQ (Ours)	**0.99**	0.97	2.67	**0.99**	**0.98**	**2.06**	**0.96**	**0.96**	**3.25**	**0.98**	**0.98**	**2.98**

ST360IQ outperformed in almost all distortion types in the CVIQ dataset and reached state-of-the-art performance in JPEG compression.

Performance Comparison(OIQA)

Table 2. Quantitative comparison of ST360IQ against the state-of-the-art on OIQA. Bold scores indicate the best performances.

		JPEG			JPEG2000			Gaussian Blur			Gaussian Noise			Overall		
	Method	PLCC↑	SRCC↑	RMSE↓	PLCC↑	SRCC↑	RMSE↓	PLCC↑	SRCC↑	RMSE↓	PLCC↑	SRCC↑	RMSE↓	PLCC↑	SRCC↑	RMSE↓
Full Reference	PSNR	0.75	0.72	1.43	0.88	0.89	1.01	0.97	0.96	1.73	0.77	0.82	1.17	0.64	0.60	1.54
	SSIM [16]	0.94	0.96	0.73	0.96	0.97	0.54	0.97	0.95	0.41	0.96	0.96	0.51	0.92	0.92	0.77
	FSIM [17]	0.95	0.96	0.65	0.95	0.95	0.63	0.97	0.96	0.40	0.96	0.96	0.48	0.93	0.93	0.72
	MS_SSIM [18]	0.97	0.94	0.49	0.94	0.92	0.74	0.88	0.87	0.79	0.82	0.84	1.04	0.70	0.68	1.42
	IW_SSIM [19]	0.95	0.96	0.67	0.97	0.97	0.44	0.87	0.84	0.84	0.91	0.92	0.73	0.76	0.75	1.31
	SR-SIM [20]	0.94	0.96	0.73	0.95	0.96	0.62	0.96	0.95	0.42	0.96	0.96	0.51	0.92	0.93	0.75
	GMSD [21]	0.95	0.94	0.93	0.95	0.93	0.65	0.90	0.85	0.74	0.85	0.88	0.96	0.77	0.76	1.27
	VSI [13]	0.95	0.97	0.65	0.96	0.96	0.56	0.97	0.96	0.38	0.96	0.96	0.94	0.93	0.93	0.72
	HaarPSI [22]	0.95	0.96	0.65	**0.98**	**0.97**	**0.36**	0.91	0.91	0.70	0.92	0.91	0.70	0.84	0.83	1.077
	LPIPS [3]	0.98	0.97	0.41	0.91	0.91	0.86	0.96	0.95	0.48	0.94	0.96	0.49	0.93	0.94	0.68
	DISTS [1]	0.98	0.99	**0.37**	0.96	0.95	0.56	**0.98**	0.95	**0.33**	0.95	0.95	0.56	0.94	0.94	0.64
	MDSI [23]	0.93	0.92	0.75	0.97	0.96	0.53	0.97	0.96	0.38	0.97	0.96	0.42	0.94	0.94	0.63
No Ref.	BRISQUE [24]	0.86	0.97	1.08	0.71	0.71	1.54	0.82	0.94	0.97	0.88	0.83	0.85	0.75	0.76	1.33
	MUSIQ [6]	0.97	0.98	0.46	0.91	0.90	0.82	0.84	0.85	0.61	0.89	0.90	0.90	0.92	0.92	0.79
	MC360IQA [8]	0.97	0.97	0.53	0.91	0.91	0.88	0.97	**0.97**	0.40	0.96	0.98	0.37	0.94	0.94	0.66
	VGCN* [9]	0.95	0.93	0.67	**0.98**	0.95	0.48	**0.98**	0.96	0.33	**0.98**	0.98	**0.35**	0.95	0.96	0.63
	VGCN+ [9]	0.89	0.88	0.98	0.92	0.89	0.90	0.90	0.85	0.78	0.96	0.94	0.47	0.88	0.89	0.92
	ST360IQ (Ours)	**0.99**	**0.99**	0.39	0.97	**0.97**	0.47	0.89	0.83	0.49	0.97	**0.99**	0.47	**0.96**	**0.97**	**0.57**

VGCN+ stands for the model trained with the same settings as our proposed method VGCN* stands for the results given in the original VGCN paper [9]

ST360IQ outperformed in all three metrics of an overall quality score predicting and gained state-of-the-art results among no-reference models.

Ablation Study

- Effect of Tangent viewports and sampling module
- Effect of model embeddings

Table 3. Contribution of using tangent viewports and saliency-guided sampling module to the final performance.

Method	CVIQ PLCC↑	CVIQ SRCC↑	OIQA PLCC↑	OIQA SRCC↑
Proposed model	0.98	0.98	0.96	0.97
w/o saliency-guided sampling	0.96	0.95	0.93	0.93
w/o tangent viewports	0.94	0.92	0.93	0.93

Table 4. Effect of using different embeddings on the performance within the proposed ST360IQ model.

Method	CVIQ PLCC↑	CVIQ SRCC↑	OIQA PLCC↑	OIQA SRCC↑
Proposed model	0.98	0.98	0.96	0.97
w/o source embed.	0.96	0.96	0.95	0.96
w/o geometric+source embed.	0.94	0.95	0.93	0.94

Conclusion

- We proposed a spherical-ViT based no-reference omnidirectional IQA method called ST360IQ
- It predicts the quality score of a 360 image by processing the image by extracting the most salient viewports, and aggregating the local quality scores estimated from them.
- Using the ViT-architecture allows us to better model the geometry of the spherical structure and the viewport biases.

REFERENCES

1. Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang, “MUSIQ: Multi-scale image quality transformer,” in ICCV, 2021, pp. 5148–5157.
2. Wei Sun, Xiongkuo Min, Guangtao Zhai, Ke Gu, Huiyu Duan, and Siwei Ma, “Mc360iqa: A multi-channel cnn for blind 360-degree image quality assessment,” IEEE J. Sel. Top. Signal Process., vol. 14, no. 1, pp. 64–77, 2020.
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in CVPR, 2016, pp. 770–778.
4. Wei Sun, Ke Gu, Siwei Ma, Wenhan Zhu, Ning Liu, and Guangtao Zhai, “A large-scale compressed 360-degree spherical image database: From subjective quality evaluation to objective model comparison,” in MMSP, 2018, pp. 1–6.
5. Huiyu Duan, Guangtao Zhai, Xiongkuo Min, Yucheng Zhu, Yi Fang, and Xiaokang Yang, “Perceptual quality assessment of omnidirectional images,” in ISCAS, 2018, pp. 1–5.
6. Jiahua Xu, Wei Zhou, and Zhibo Chen, “Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks,” IEEE Trans. Circuits Syst. Video Technol., 2020