Semantically Adaptive JND Modeling with Object-wise Feature Characterization and Cross-object Interaction

Xia Wang[*], Haibing Yin[*], Tingyu Hu[*], and Qinghua Sheng[†]

[*]School of Communication Engineering
Hangzhou Dianzi University
Hangzhou, 310018, PR China
{xiawang,yhb,tingyuhu}@hdu.edu.cn

[†]School of Electronics and Information
Hangzhou Dianzi University
Hangzhou, 310018, PR China
sheng7@hdu.edu.cn

This work proposed a spatio-temporal JND model based on semantic attention. Firstly, the principal semantic features affecting visual attention are extracted, including the semantic sensitivity, objective area and shape, central bias and contextual complexity, and the HVS responses of these four features are explored and quantified. Secondly, the semantic attention model is constructed by inscribing the attentional competition model, considering the interaction between different objects with limited perception resources. Finally, the obtained semantic attention weighting factor is combined with the basic spatial attention model to develop an improved transform domain JND model. Detailed performance results of different JND models are shown in Tab. 1. The simulation results validate that the proposed JND profile is highly consistent with HVS, with strong competitiveness among the state-of-the-art models.

Table 1: Performance Comparison of Different JND Models.

Sequences	Bae2017 [1]		Zeng2019 [2]		Xing2021 [3]		Propose	
	PSNR	DMOS	PSNR	DMOS	PSNR	DMOS	PSNR	DMOS
BasketballDrill	26.73	14.24	32.10	22.18	26.19	11.59	**25.22**	**10.88**
PartyScene	29.33	10.18	31.05	23.88	27.94	12.24	**27.10**	**10.06**
RaceHorses	31.41	9.59	31.21	22.76	**29.50**	10.47	29.96	**8.71**
FourPeople	26.86	14.41	30.06	24.41	25.90	14.88	**25.30**	**11.94**
Johnny	27.76	13.65	31.34	21.76	28.98	11.71	**26.91**	**7.29**
KristenAndSara	28.39	12.47	30.08	24.12	24.94	12.94	**24.56**	**7.71**
Kimono1	27.54	13.94	28.96	25.06	26.55	14.76	**25.94**	**7.41**
ParkScene	26.61	13.00	29.07	23.18	24.94	13.12	**24.56**	**8.18**
BasketballDrive	31.11	10.76	32.15	22.47	30.52	10.65	**29.15**	**9.88**
BQTerrace	28.58	10.53	30.39	23.65	29.73	12.71	**28.54**	**8.12**
Average	**28.43**	**12.49**	**30.64**	**23.47**	**27.83**	**12.51**	**27.03**	**9.42**

References

[1] S. H. Bae and M. Kim, "A dct-based total jnd profile for spatiotemporal and foveated masking effects," *IEEE Trans. Circuits Syst. Video Technol*, vol. 27, no. 6, pp. 1196–1207, June 2017.

[2] Z. Zeng, H. Zeng, J. Chen, J. Zhu, Y. Zhang, and K. K. Ma, "Visual attention guided pixel-wise just noticeable difference model," *IEEE Access*, vol. 7, pp. 132111–132119, September 2019.

[3] Y. Xing, H. Yin, Y. Zhou, Y. Chen, and C. Yan, "Spatiotemporal just noticeable difference modeling with heterogeneous temporal visual features," *Displays*, vol. 70, pp. 102096, October 2021.