ByteDance



MSTRIQ: No Reference Image Quality Assessment Based on Swin Transformer with Multi-Stage Fusion

Jing Wang, Haotian Fan, Xiaoxia Hou, Yitian Xu, Tao Li, Xuechao Lu, Lean Fu ByteDance Inc

PROPOSED METHOD



Learning Objectives

 \succ The predicted results of the IQA model should be close to human ratings

Regression loss:

evaluates the absolute distance between the MOSs and the predicted results

$$loss_{reg} = \frac{1}{2N} \sum_{i=1:N} |\hat{y}_i - y_i|^2$$

Exponential rank loss:

explicitly exploits relative rankings of image pairs in the dataset

$$pss_{rank} = \frac{2}{N} \sum_{i=0:2:N} \begin{cases} e^{\hat{y}^{i} - \hat{y}^{i+1}}, & \text{if } y^{i} < y^{i-1} \\ 0, & \text{others} \end{cases}$$

Data Augmentation

- > Training: augmentation techniques are all used with specific probabilities for Rotation and Color Space Changing respectively.
- Testing: only Resize and Crop are used.





Weighted sampling

up-sampling minority classes

Datasets	KonIQ-10k		Methods	PLCC	SRCC
Methods	PLCC	SRCC	mse	0.951	0.940
Ours without feature fusion	0.953	0.942	rank	0.941	0.945
Ours with feature fusion	0.954	0.946	mse+rank	0.954	0.946

EXPERIMENT RESULTS

Evaluation Metrics

<u>PLCC</u>: Pearson product-moment correlation coefficient <u>SROCC</u>: Spearman rank-order correlation coefficient MainScore: PLCC + SROCC

Results on Public Dataset				Results on PIPAL Dataset				
Datasets	KonIQ-1	l0k	TID201	3	IQA Name	MainScore	SRCC	PLCC
Methods	PLCC	SRCC	PLCC	SRCC	PSNR	0.572	0.269	0.303
BRISQUE	0.681	0.665	0.610	0.544	SSIM	0.785	0.377	0.407
HyperIQA	0.917	0.906	0.858	0.840	LPIPS-Alex	1.176	0.584	0.592
MetaIQA	0.887	0.850	0.868	0.856	FSIM	1.138	0.528	0.610
FPR	0.901	0.899	0.887	0.872	NIQE	0.142	0.030	0.112
TReS	0.928	0.915	0.883	0.863	MA	0.398	0.174	0.224
Ours	0.954	0.946	0.895	0.882	PI	0.276	0.123	0.153
					Brisque	0.184	0.087	0.097
Scalar Plots with MOSs			Ours	1.437	0.737	0.700		
TID2013KonIQ 10k				Results of NTIRE NR-IQA Challenge				
		Teams	MainScore	PLCC	SRCC			
		1st	1.444	0.740	0.704			
			DTIQA	1.437	0.737	0.700		
			3rd	1.422	0.725	0.697		
Z model pred	4 0	t mo	del predict	ed result	4th	1.407	0.726	0.681
model pred	aleteu resul				5th	1.390	0.720	0.671



[1] Z. Liu, Y. Lin, Y. Cao, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE Int. Conf. Comput. Vis., 10012-10022, 2021





ABLATIONS

REFERENCES