Image Quality Assessment with Gradient Siamese Network

網易 NetEase

Heng Cong* Lingzhi Fu* Rongyu Zhang* Yusheng Zhang

Hao Wang Jiarong He Jin Gao

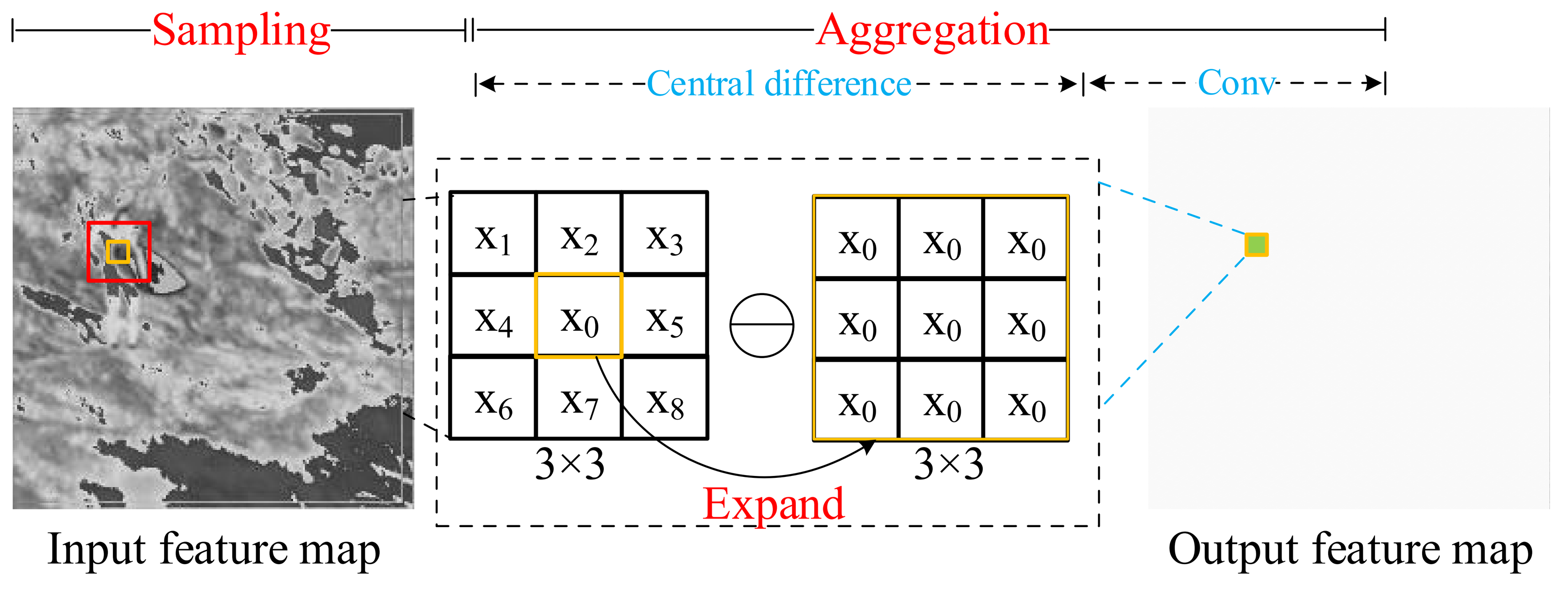
CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Introduction

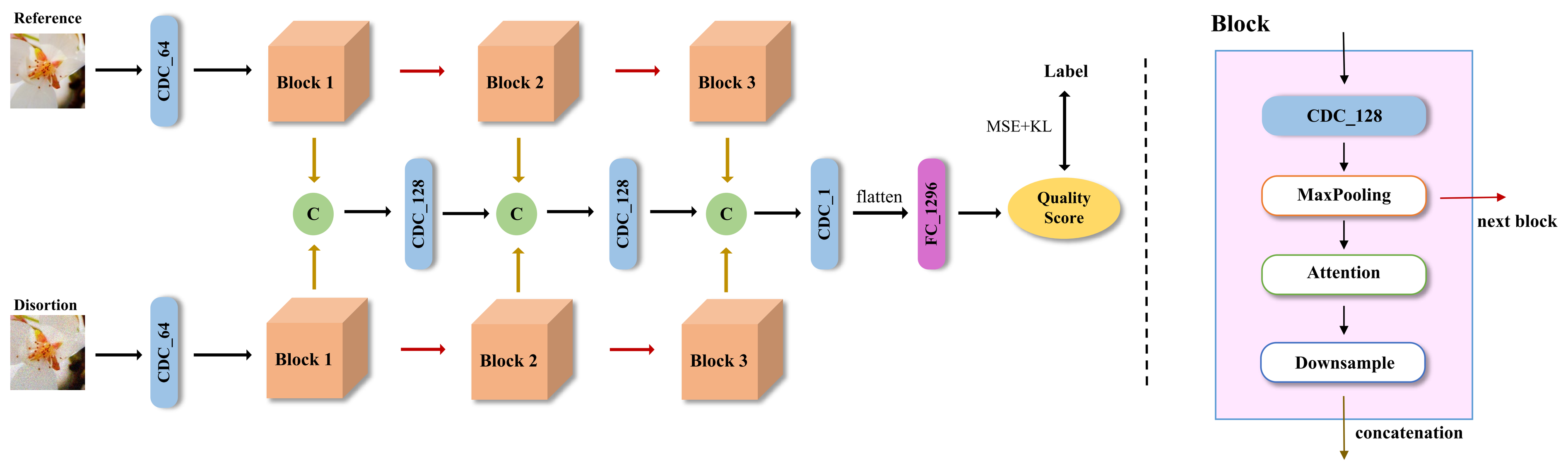
- The method of extracting image gradient information by CNN is effective but ignores the key fine-grained information.
- Traditional regression networks use an MSE function to supervise the network output, but do not take into account the logical relationship of image quality score ranking.
- Our work focuses on efficiently extracting contrast gradient information of image pairs in the full reference problem.

Proposed Method

- Combined with ***Central Difference Convolution (CDC)***, which can efficiently extract gradient features and semantic features.



- ***Gradient Siamese Network (GSN)***, based on CDC and spatial attention. A multi-level feature fusion module that concatenates features at the same level and then fuses features from different levels.



Loss Function

- We successfully apply KL divergence loss to image quality assessment task for raising the importance of relative order learning.

Normalize Output

$$\hat{S}_i = \frac{\left(\sum_{i=1}^{N}\left|\hat{Q}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{Q}_i\right|^q\right)^{\frac{1}{q}}}{\hat{Q}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{Q}_i}$$

$$S_i = \frac{\left(\sum_{i=1}^{N}\left|Q_i - \frac{1}{N}\sum_{i=1}^{N}Q_i\right|^q\right)^{\frac{1}{q}}}{Q_i - \frac{1}{N}\sum_{i=1}^{N}Q_i}$$

$$\widehat{W}_i = \frac{exp\hat{Q}_i}{\sum_{i=1}^{N} exp\hat{Q}_i}$$

$$W_i = \frac{expQ_i}{\sum_{i=1}^{N} expQ_i}$$

Loss

$$L_{pair}(Q,\hat{Q}) = MSE(\hat{S}_i - S_i)$$

$$L_{list}(Q,\hat{Q}) = \sum_{i=1}^{N}\widehat{W}_i \times log\frac{\widehat{W}_i}{W_i}$$

$$L(Q,\hat{Q}) = \alpha \cdot L_{pair}(Q,\hat{Q}) + \beta \cdot L_{list}(Q,\hat{Q})$$

Public Comparison

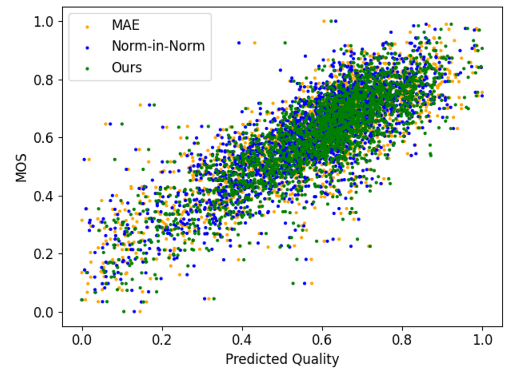
- For CSIQ and TID2013, our proposed GSN is more robust and ranked in the top two in terms of main score.

Method	CSIQ				TID2013			
	PLCC	SRCC	KRCC	MS	PLCC	SRCC	KRCC	MS
PSNR [13]	0.819	0.81	0.601	1.629	0.677	0.687	0.496	1.364
SSIM [14]	0.852	0.865	0.68	1.717	0.777	0.727	0.545	1.504
MS-SSIM [24]	0.889	0.906	0.730	1.795	0.830	0.786	0.605	1.616
VSI [37]	0.928	0.942	0.786	1.870	0.900	0.897	0.718	**1.797**
MAD [25]	0.95	0.947	0.797	**1.897**	0.827	0.781	0.604	1.608
VIF [38]	0.913	0.911	0.743	1.824	0.771	0.677	0.518	1.448
FSIM [39]	0.919	0.931	0.769	1.850	0.877	0.851	0.667	1.728
NLPD [40]	0.923	0.932	0.769	1.855	0.839	0.800	0.625	1.639
GMSD [41]	0.945	0.950	0.804	1.895	0.855	0.804	0.634	1.659
DeepIQA [27]	0.901	0.909	0.732	1.810	0.834	0.831	0.631	1.665
PieAPP [42]	0.877	0.892	0.715	1.769	0.859	0.876	0.683	1.735
LPIPS [43]	0.896	0.876	0.689	1.772	0.749	0.67	0.497	1.419
DISTS [44]	0.928	0.929	0.767	1.857	0.855	0.830	0.639	1.685
GSN(ours)	0.944	0.951	0.801	**1.896**	0.892	0.885	0.699	**1.776**

Ablation Studies

- All models using CDC can achieve an MS value over 1.6, which demonstrates that CDC is beneficial and effective in IQA.
- The MS score of M2 (CNN & MSE+KL) is 0.053 higher than the MS score of M1 (CNN & MSE) by adding KL function.

Model	Conv.Type		Loss		PIPAL [11]	
	CDC	CNN	MSE	MSE+KL	PLCC/SRCC	MS
M1		√	√		0.785/0.751	1.536
M2		√		√	0.802/0.787	1.589
M3	√		√		0.828/0.804	1.632
M4	√			√	**0.835/0.811**	**1.646**

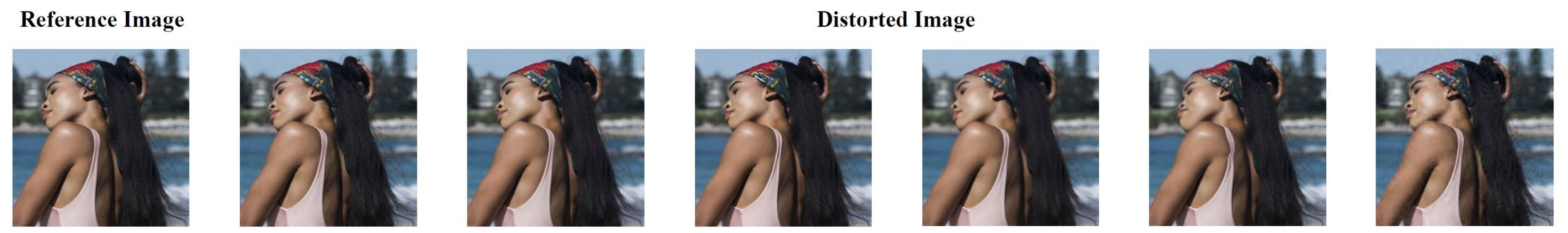


NTIRE 2022 IQA-FR Challenge

- GSN won the second place in NTIRE track 1 Full-Reference!

Team	mainScore	PLCC	SRCC
1st	1.651	0.828	0.822
GSN(ours)	**1.642**	**0.827**	**0.815**
3rd	1.640	0.823	0.817
4th	1.541	0.775	0.766
5th	1.538	0.772	0.765
6th	1.501	0.763	0.737
7th	1.450	0.763	0.737
8th	1.403	0.703	0.701

- Visual result from the validation set of the NTIRE 2022 challenge.



MOS ↑	1554.74(1)	1554.65(2)	1553.11(3)	1549.99(4)	1412.58(5)	1381.01(6)
PSNR ↑	36.40(1)	35.95(2)	30.45(5)	32.76(3)	27.45(6)	30.63(4)
SSIM ↑	0.905(4)	0.949(1)	0.889(5)	0.907(3)	0.925(2)	0.770(6)
MS-SSIM ↑	0.985(2)	0.993(1)	0.981(4)	0.985(3)	0.956(5)	0.946(6)
LPIPS ↓	0.020(1)	0.076(5)	0.051(3)	0.066(4)	0.037(2)	0.079(6)
GSN(ours) ↑	0.616(1)	0.610(2)	0.591(3)	0.582(4)	0.528(6)	0.530(5)

Conclusion

- Central difference convolution is introduced to GSN for the image fine-grained.
- A fusion structure is incorporated which handles multi-level features.
- KL divergence loss helps improve prediction accuracy.