LiveVideo StackCon 音视频技术大会

技术开启新“视”界
Technology Bring New Vision

无参考图像视频质量评价

蒋婷婷
北京大学数字媒体所

LiveVideoStack
音视频技术社区

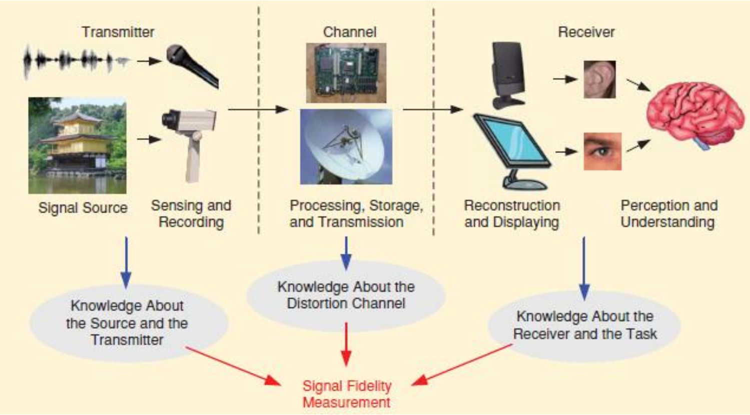
CSDN

1 图像质量评价简介

2 无参考图像质量评价研究进展

图像质量评价

- 图像质量评价概述
- 评价方法
- 人工效应分析



图像质量的含义

- 对一幅图像视觉感受的主观评价
- 图像质量的理论基础
  - 图像不是**简单的二维信号**而是作为**视觉信息的载体**
  - 视觉感知过程不是**信号**处理过程而应该是**信息**处理过程
  - 视觉感知过程不是**独立**的过程，而是人们与环境**交互**的基本阶段
  - 图像质量不是指**图像失真的可见性**，而是指在视觉交互过程中**对于输入信息的感知度**
- 图像质量评价的含义
  - **逼真度**：描述被评价图像与标准图像的偏离程度
  - **可懂度**：表示图像能向人或计算机提供信息的能力

图像质量评价的应用

- 对采集的图像进行质量评价，以此判断采集设备和成像系统的性能优劣
- 经过通信传输后在用户端或终端评价图像的质量需要无参考的质量评估，以此来判断通讯传输技术的优劣
- 在数字图像处理过程中，需要分别对源图像及失真图像进行质量评价，以此来判断算法的优劣
- 图像压缩

质量评价方法分类

- **主观评价：**观察者评分来判断图像质量
  - 优点：准确
  - 缺点：无法用数学模型进行描述，费时费力
- **客观评价**
  - **参考源可用性：**全参考、无参考、部分参考质量评价
  - **评价处理方式：**空间域、频域、空间域和频域综合
  - **评价指标角度：**单因素(噪声、模糊、块效应等)、综合因素质量评价方法
  - **视觉心理生理角度：**自顶向下的方法、结合人眼视觉系统的自底向上的方法
  - **应用智能角度：**基于神经网络、机器学习、模糊理论、贝叶斯理论等

性能指标和评价准则

- VQEG(**Video Quality Experts Group)**组织提出质量模型评价标准包括：
  - 预测的精确性
  - 预测的单调性
  - 预测的一致性
- 主观打分: Mean Opinion Score (MOS)
- 客观评价：Quality Rating (QR)
- 将QR向MOS进行非线性拟合成MOS_p
- 比较MOS和MOS_p

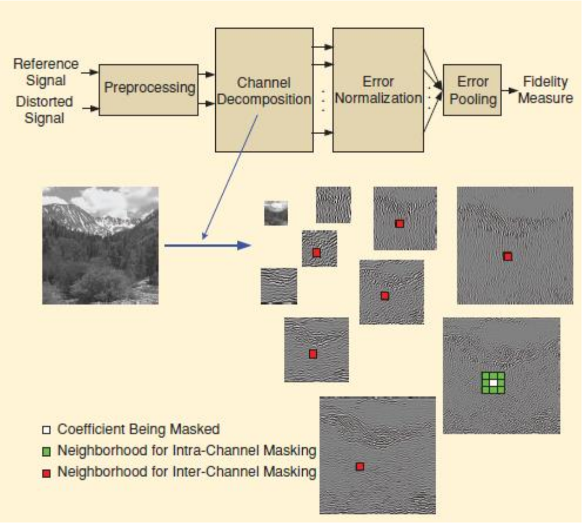
主观质量评价

- 国际电信联盟：标准化工作
- 度量尺度：
  - **绝对尺度**:将图像直接按照视觉感受分级评分
  - **相对尺度**:在一组图像中，按该组图像的相对优劣进行分级
- 存在的问题：
  - 缺乏稳定性，不能保证评价的可重复性
  - 无法应用数学模型对其进行描述，费时费力
  - 无法实现嵌入式/实时处理，不适用于工程化

客观质量评价

根据参考源的可用性分为：

- 全参考
- 部分参考
- 无参考

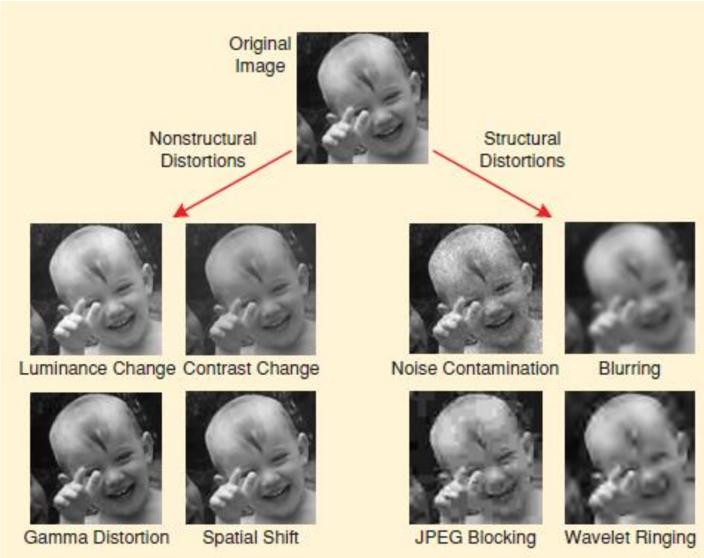


全参考质量评价（1）

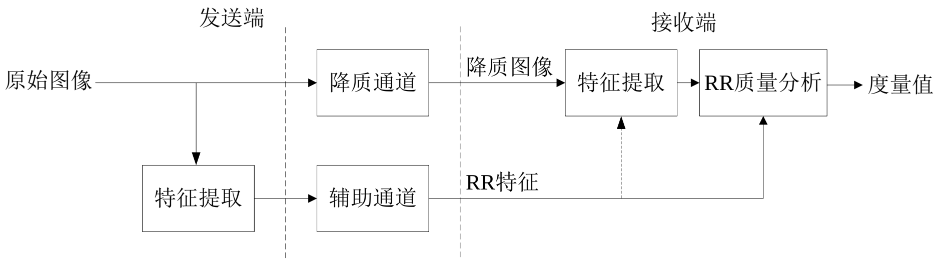
- **基于全像素失真统计的传统评价方法**
  - PSNR、MSE、MAE、RMS、SNR、STD
- **基于人眼视觉系统（HVS）的评价方法**
  - 基于视觉感知的算法模型[Chou 95][Winkler 99]
  - 基于视觉兴趣加权的算法模型[Chiu 96]
- **基于图像理解的评价方法**
  - 分层模型[Hamada 99]
    - 噪声层、纹理层、目标层
  - 分割模型[Pessoa98]
    - 平坦区、纹理区、边缘区

全参考质量评价（2）

- 基于图像结构相似性的评价方法 SSIM
- 基于学习的方法



部分参考质量评价

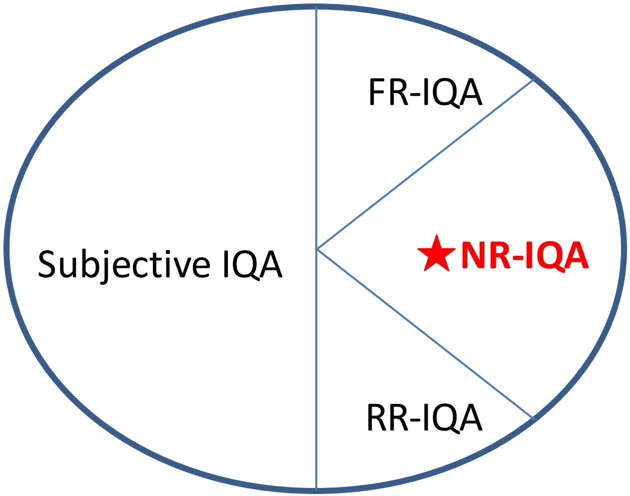


- 基于源数据的信息提取方法
  - 提取图像源本身的特征信息作为质量评价依据
  - 非期望特征提取算法（反映损伤程度的特征）
  - 期望特征提取算法
- 基于非源数据的信息添加方法
  - 在发送端(或编码端) 添加非原始图像数据的额外信息
  - 在接收端通过分析这些信息的损耗程度，侧面反映图像质量

无参考质量评价方法

- 针对噪声的评价
- 针对人工效应的评价
  - 图像压缩技术引入人工效应：块效应、模糊效应、振铃效应
- 其它评价方法
  - 基于自然图像统计规律的方法
  - 基于学习的方法

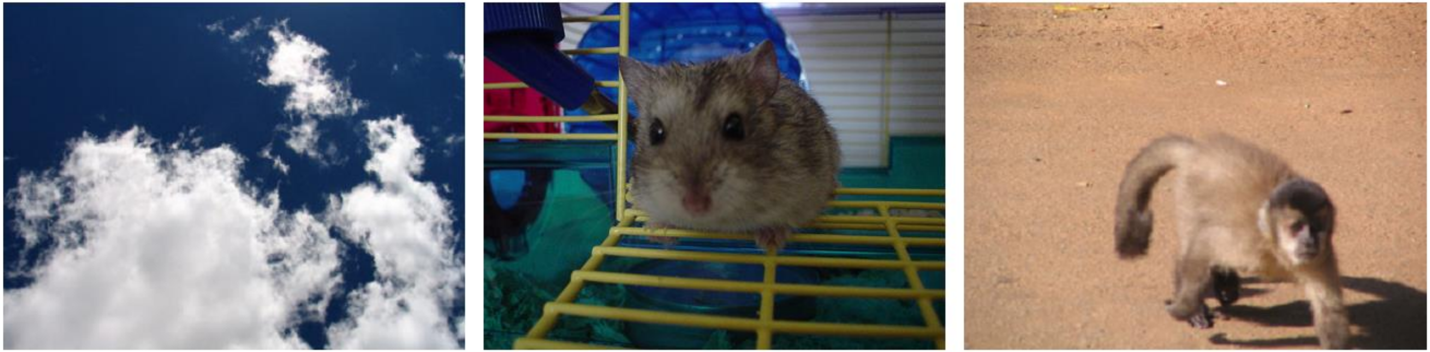
Background of image quality assessment (IQA)



The effect of image content variation on IQA

Traditional methods tend to overestimate the image quality in complex contents but underestimate it in simple contents.

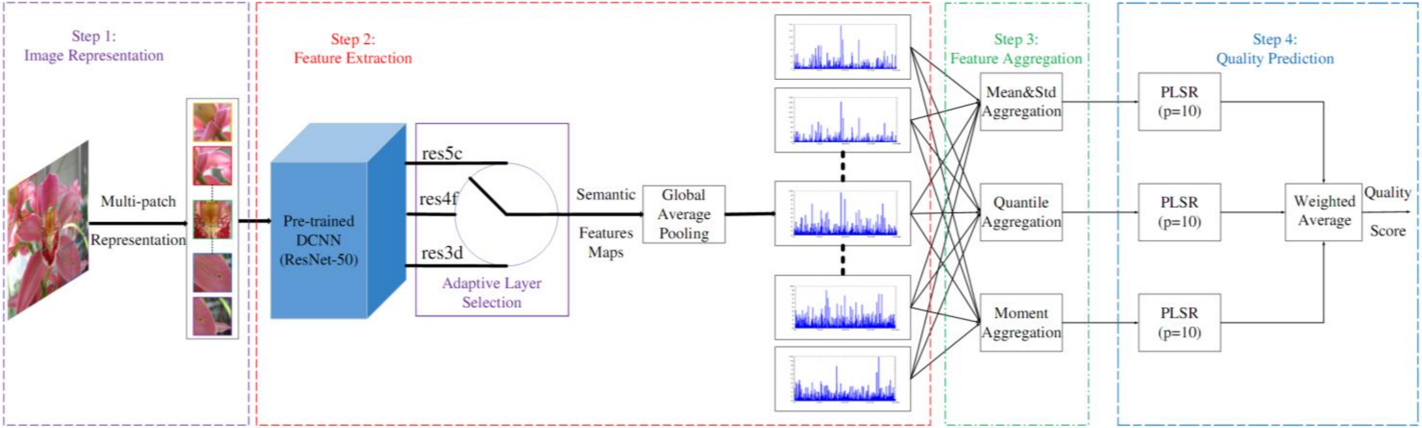
The effect of image content variation on IQA



(a) The clear blue sky (b) A blurry mouse (c) A blurry monkey

Which Has Better Visual Quality: The Clear Blue Sky or a Blurry Animal?

Reducing the effect by image-content-aware features



Dingquan Li, Tingting Jiang, Weisi Lin, and Ming Jiang. Which Has Better Visual Quality: The Clear Blue Sky or a Blurry Animal?. IEEE Transactions on Multimedia, 2018, accepted.

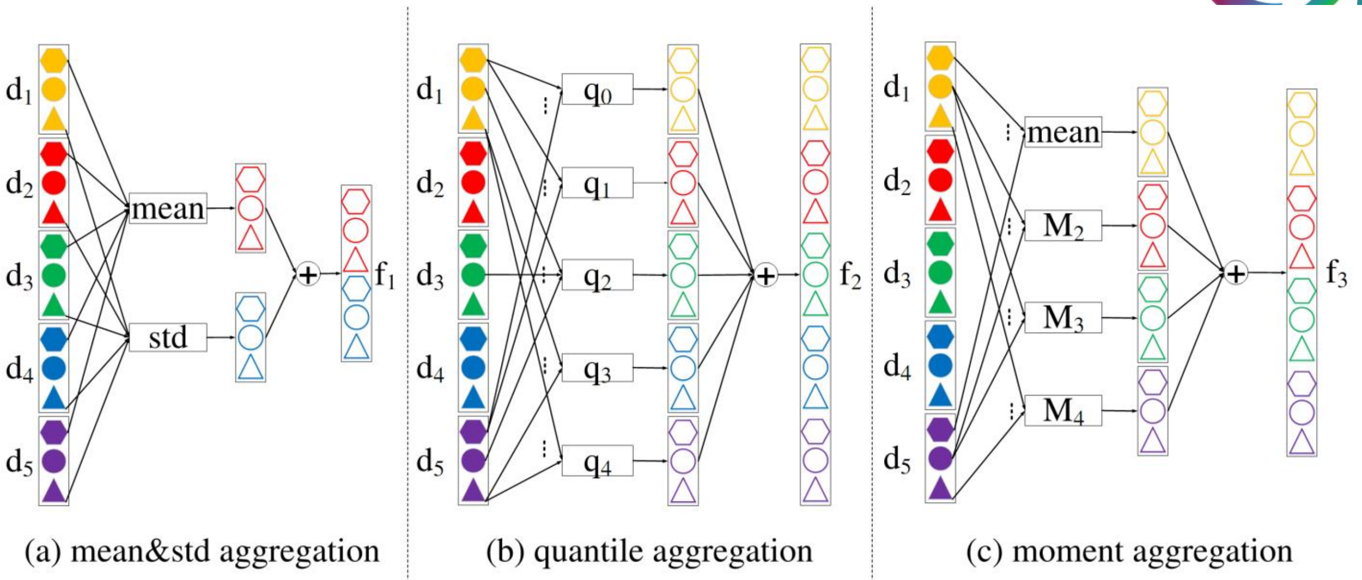


Fig. 6. [Best viewed in color.] An illustration of the three statistical structures used for feature aggregation. The inputs are $n = 5$ features $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_5\}$, and the feature dimension is $l = 3$. $(\mathbf{q}_0, \mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4)$ indicates the 5 quartiles, and $\mathbf{M}_r^r$ equals the central moment of order r $(r = 2, 3, 4)$. For clarity, some links between patch features and statistical functions are omitted.

Quantifying the Effect of Image Content Variation

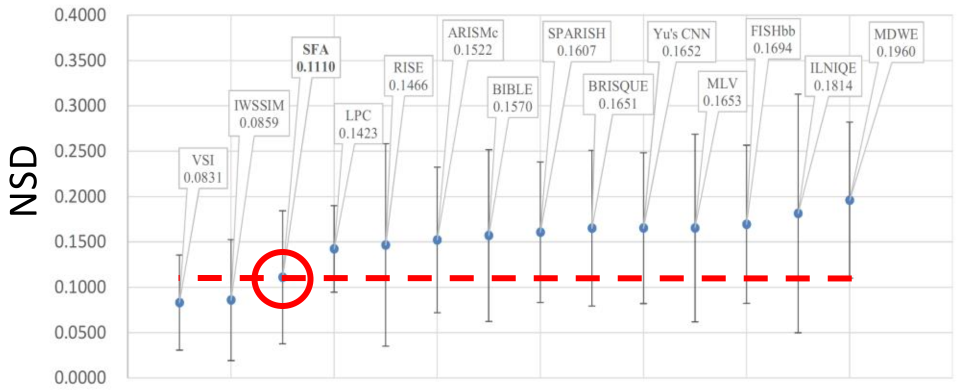
- Quality-indiscriminate image pair

Symbol	Meaning
S	a quality-indiscriminate image dataset including N images
$\mathrm{std}_o(S)$	the standard deviation of the objective scores on the dataset S.
$\mathrm{std}_s(S)$	the standard deviation of the subjective scores on the dataset S.
R	the range of subjective quality scores
$[x]_+$	the positive part of x
NSD	The measure for quantifying the effect of image content variation

$$\mathrm{NSD} = \frac{[\mathrm{std}_o(\mathcal{S}) - 2\mathrm{std}_s(\mathcal{S})]_+}{R/2\sqrt{3}}$$

Quantitative Results

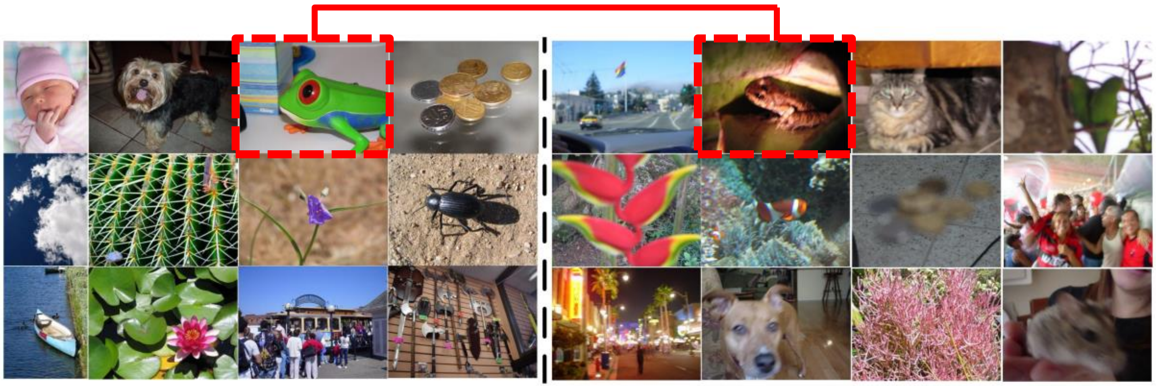
- Quality-indiscriminate image pair



NSDs Deviate from 0.

Quantifying the Effect of Image Content Variation

- Quality-discriminable image pair



22,792 (154 × 148) quality-discriminable image pairs

Quantitative Results

- Quality-discriminable image pair

Category	Method	Accuracy
Learning-free	MDWE [25]	73.61%
	MLV [27]	73.21%
	ARISMc [26]	51.68%
	FISHbb [29]	84.49%
	LPC [24]	73.26%
	BIBLE [31]	73.01%
Learning-based	SPARISH [38]	76.54%
	RISE [37]	76.65%
	Yu's CNN [46]	75.95%
	BRISQUE [8]	58.17%
	ILNIQE [9]	85.01%
	SFA (Proposed)	**96.87%**

More than 11000 failure cases

More than 3400 failure cases

Experiments and analysis

- Databases
- Performance Comparison
- Generalization Capability

Databases

TABLE II
GENERAL INFORMATION OF THE SEVEN DATABASES USED FOR THE COMPARATIVE EXPERIMENTS. "# REFERENCE IMAGE" MEANS THE NUMBER OF REFERENCE IMAGES, AND "# BLUR IMAGE" MEANS THE NUMBER OF BLUR IMAGES.

Database	# Reference image	# Blur image	Blur type	Score type*	Score range
LIVE [14]	29	145	Gaussian blur	DMOS	[0 100]
TID2008 [15]	25	100	Gaussian blur	MOS	[0 9]
TID2013 [16]	25	125	Gaussian blur	MOS	[0 9]
MLIVE1 [17]	15	225	Gaussian blur with white Gaussian noise	DMOS	[0 100]
MLIVE2 [17]	15	225	Gaussian blur with JPEG compression	DMOS	[0 100]
BID [18]	-	586	Realistic blur (out-of-focus, motion, *etc.*)	MOS	[0 5]
CLIVE [19]	-	1162	Realistic blur	MOS	[0 100]

* DMOS indicates the difference of mean opinion scores (MOS) between the test image and its reference image.

Performance Comparison

Method	LIVE [14]					TID2008 [15]					TID2013 [16]				
	SROCC↑	KROCC↑	PLCC↑	RMSE↓	OR↓	SROCC	KROCC	PLCC	RMSE	OR	SROCC	KROCC	PLCC	RMSE	OR
IWSSIM [2]	0.9723	0.8733	0.9698	4.4734	40.00%	0.9680	0.8707	0.9533	0.3459	45.00%	0.9723	0.8787	0.9526	0.3753	56.80%
VSI [3]	0.9538	0.8300	0.9535	5.4508	48.00%	0.9592	0.8496	0.9551	0.3397	50.00%	0.9669	0.8581	0.9571	0.3593	56.00%
MDWE [25]	0.9188	0.7800	0.9377	6.4427	52.00%	0.8556	0.6579	0.8660	0.5697	70.00%	0.8466	0.6467	0.8698	0.6039	72.00%
MLV [27]	0.9431	0.8133	0.9578	5.2170	48.00%	0.8977	0.7158	0.9075	0.4837	65.00%	0.9142	0.7446	0.9226	0.4762	64.00%
ARISMc [26]	0.9585	0.8467	0.9684	4.6117	**40.00%**	0.8851	0.7124	0.8872	0.5266	65.00%	0.9108	0.7513	0.9149	0.4938	64.00%
FISHbb [29]	0.9469	0.8267	0.9570	5.2410	48.00%	0.8737	0.6807	0.8916	0.5160	65.00%	0.8900	0.7067	0.9087	0.5100	68.00%
LPC [24]	0.9469	0.8133	0.9326	6.6480	56.00%	0.8805	0.6860	0.8858	0.5334	65.00%	0.9049	0.7267	0.9086	0.5132	64.00%
BIBLE [31]	**0.9638**	0.8533	0.9711	**4.3871**	**40.00%**	0.9114	0.7441	0.9178	0.4575	**60.00%**	0.9131	0.7446	0.9264	0.4615	64.00%
SPARISH [38]	**0.9638**	**0.8600**	0.9693	4.4870	**40.00%**	0.9126	0.7474	0.9164	0.4628	**60.00%**	0.9102	0.7400	0.9228	0.4716	64.00%
RISE [37]	0.9492	0.8267	0.9594	5.6563	48.00%	0.9203	0.7757	0.9235	0.4891	**60.00%**	0.9300	0.7800	0.9342	0.4971	68.00%
Yu's CNN [46]	0.9469	0.8200	0.9486	6.5674	48.00%	0.8752	0.6737	0.8784	0.6426	70.00%	0.8929	0.7067	0.9020	0.6195	76.00%
BRISQUE [8]	-	-	-	-	-	0.8782	0.6947	0.8865	0.5330	65.00%	0.8878	0.7067	0.8963	0.5536	68.00%
ILNIQE [9]	0.9308	0.7933	0.9444	6.1241	56.00%	0.8451	0.6491	0.8617	0.5782	70.00%	0.8466	0.6533	0.8675	0.6134	76.00%
SFA (Proposed)	0.9631	**0.8600**	**0.9722**	4.7469	**40.00%**	**0.9368**	**0.8000**	**0.9455**	**0.4193**	**60.00%**	**0.9477**	**0.8180**	**0.9542**	**0.4281**	**60.00%**

BRISQUE is trained on the full LIVE IQA database.

Performance Comparison

Method	MLIVE1 [17] SROCC	KROCC	PLCC	RMSE	OR	MLIVE2 [17] SROCC	KROCC	PLCC	RMSE	OR
IWSSIM [2]	0.9198	0.7624	0.9340	6.4245	0.00%	0.9103	0.7495	0.9386	6.4895	0.00%
VSI [3]	0.8882	0.7179	0.9104	7.5412	0.00%	0.8797	0.7067	0.9131	7.6284	0.00%
MDWE [25]	0.0869	0.0607	0.2447	17.9239	6.67%	0.5632	0.4107	0.6465	14.4053	2.22%
MLV [27]	0.4687	0.3175	0.6422	13.9836	4.44%	0.8256	0.6202	0.8827	8.9481	**0.00%**
ARISMc [26]	-0.2926	-0.2116	0.3960	17.0197	8.89%	0.8763	0.7125	0.9214	7.3130	**0.00%**
FISHbb [29]	0.3087	0.2114	0.2996	16.7142	6.67%	0.7598	0.5642	0.8560	9.7748	**0.00%**
LPC [24]	0.4401	0.3074	0.6585	13.7785	4.44%	0.7018	0.5023	0.8441	10.1885	**0.00%**
BIBLE [31]	0.1563	0.0971	0.3147	17.4678	8.89%	0.8337	0.6384	0.8953	8.2416	**0.00%**
SPARISH [38]	-0.0532	-0.0313	0.3370	17.3901	6.67%	0.9132	0.7556	0.9413	6.4184	**0.00%**
RISE [37]	0.8613	0.6761	0.8877	10.4500	**0.00%**	0.8846	0.7152	0.9240	8.6906	**0.00%**
Yu's CNN [46]	0.8828	0.7125	0.8959	10.4125	**0.00%**	0.8759	0.7040	0.9140	9.0764	**0.00%**
BRISQUE [8]	0.3055	0.2239	0.4071	16.8893	8.89%	0.8200	0.6458	0.9006	8.1076	**0.00%**
ILNIQE [9]	0.9219	0.7652	0.9290	**6.7615**	**0.00%**	0.9104	0.7495	0.9278	**7.1369**	**0.00%**
SFA (Proposed)	**0.9373**	**0.7899**	**0.9419**	7.5586	**0.00%**	**0.9404**	**0.8000**	**0.9468**	7.4790	**0.00%**

Performance Comparison

Method	BID [18] SROCC	KROCC	PLCC	RMSE	OR	CLIVE [19] SROCC	KROCC	PLCC	RMSE	OR
MDWE [25]	0.3067	0.2123	0.3538	1.1639	23.08%	0.4313	0.2956	0.4988	17.5025	6.90%
MLV [27]	0.3169	0.2199	0.3750	1.1561	22.22%	0.3412	0.2318	0.4076	18.4350	7.76%
ARISMc [26]	-0.0151	-0.0105	0.1929	1.2245	26.50%	0.2427	0.1631	0.3554	18.8947	8.19%
FISHbb [29]	0.4736	0.3254	0.4853	1.0894	18.80%	0.4865	0.3320	0.5380	17.0310	6.47%
LPC [24]	0.3150	0.2159	0.4053	1.1408	22.22%	0.1483	0.0968	0.3490	18.9205	7.76%
BIBLE [31]	0.3609	0.2449	0.3923	1.1469	22.22%	0.4260	0.2931	0.5178	17.3007	6.90%
SPARISH [38]	0.3074	0.2088	0.3555	1.1659	23.08%	0.4015	0.2750	0.4843	17.6702	7.33%
RISE [37]	0.5632	0.3978	0.5681	1.0543	17.09%	0.5152	0.3586	0.5550	17.1360	6.03%
Yu's CNN [46]	0.5572	0.3902	0.5600	1.0649	20.51%	0.5017	0.3491	0.5010	18.3058	8.19%
BRISQUE [8]	0.1051	0.0678	0.2246	1.2166	26.50%	0.3153	0.2136	0.3758	18.7053	8.62%
ILNIQE [9]	0.4963	0.3439	0.5192	1.0649	17.95%	0.4401	0.3013	0.5102	17.3930	6.47%
SFA (Proposed)	**0.8263**	**0.6334**	**0.8399**	**0.6859**	**5.98%**	**0.8119**	**0.6195**	**0.8331**	**11.3525**	**0.86%**

Statistical significance test

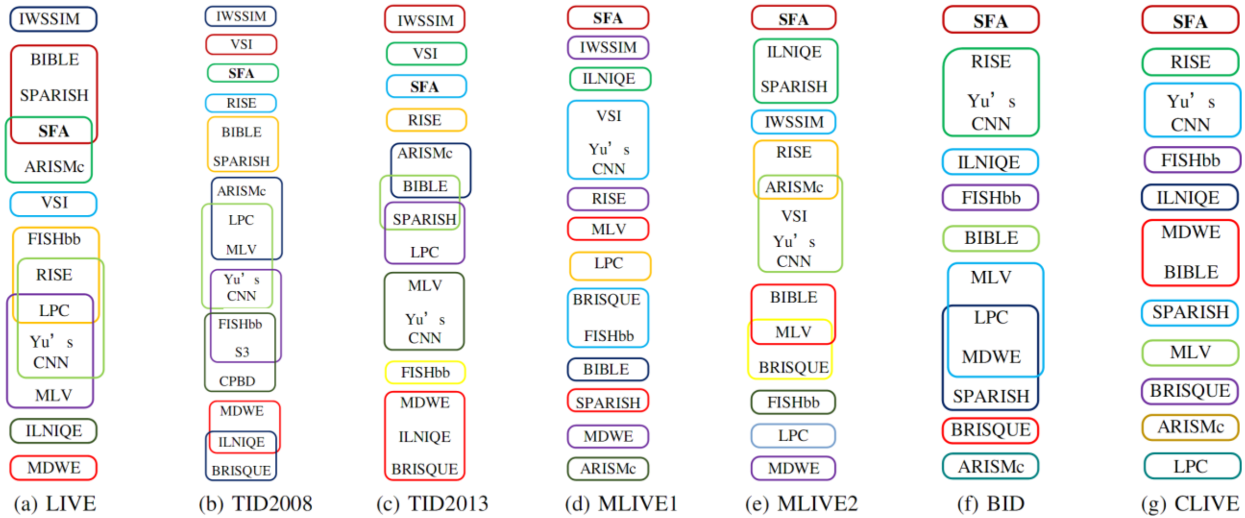


Fig. 8. [Best viewed in color.] Global ranking and grouping of methods by their statistical significance results. The methods on the upper positions achieve a better performance, and the methods within the same rectangle are statistically indistinguishable, *i.e.*, their performances are similar.

Generalization Capability

Train \ Test	LIVE	TID2008	TID2013	MLIVE1	MLIVE2	BID	CLIVE
LIVE	0.9631/0.9492	**0.9313**/0.9138	**0.9460**/0.9339	**0.3732**/0.1823	**0.7168**/0.6192	**0.5267**/0.0080	**0.4972**/0.2857
TID2008	**0.9429**/0.8638	0.9368/0.9203	**0.9815**/0.8696	**0.3597**/0.0483	**0.6834**/0.6029	**0.3667**/0.1506	**0.4664**/0.0638
TID2013	**0.9165**/0.8497	**0.9839**/0.8913	0.9477/0.9300	**0.2801**/0.3383	**0.6191**/0.4543	**0.2769**/0.0900	**0.4832**/0.2317
MLIVE1	0.8534/**0.8603**	**0.8161**/0.7775	**0.7922**/0.7157	0.9373/0.8613	**0.9025**/0.6868	**0.4474**/0.3896	0.2036/**0.2334**
MLIVE2	**0.9007**/0.7926	**0.8570**/0.8056	**0.8394**/0.6544	**0.7917**/0.4859	0.9404/0.8846	**0.4609**/0.2261	**0.3682**/0834
BID	0.7945/**0.8760**	0.7600/**0.8017**	**0.7602**/0.7106	**0.7570**/0.5504	**0.8129**/7607	0.8263/0.5632	**0.6362**/0.1931
CLIVE	**0.8897**/0.8156	**0.8603**/0.7791	**0.8796**/0.7255	**0.5643**/0.0672	**0.7995**/4754	**0.7380**/0.3613	0.8119/0.5152

Fig. 9. [Viewed in color.] The SROCC values in the form of SFA/RISE in the cross-database evaluation. In each entry, the better value is indicated in bold. Note that the intra-database experimental results are also shown (in gray) as a reference. The numerical values in red mean that the corresponding SROCC values are negative. The blue blocks emphasize the results whereby both training and testing data are simulated/realistic blur.

Conclusion

- An analysis of the impact of image content variation on NR-IQA methods verifies that deep semantic features can alleviate this impact.
- A novel NR-IQA framework is proposed based on semantic feature aggregation
- Comprehensive experiments verifies the superiority and generalization capability of the proposed method.

Thank you

LiveVideoStack
音视频技术社区

CSDN