

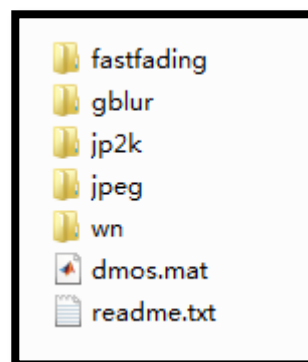
Report of Image Quality Assessment

现有图像质量评价算法共 11 种，将上述算法在 LIVE 数据库上用 Matlab 运行的结果及分析为这篇报告的主要内容。

1 Introduction

1.1 LIVE database

LIVE (Laboratory for image & video engineering) 图像质量评价数据库由美国德克萨斯大学建立，是应用最为广泛的一个共享数据库。库中含有 29 幅参考图像，以及由这些图像经过一定的降质生成的 5 种不同类型失真图像共计 779 个。

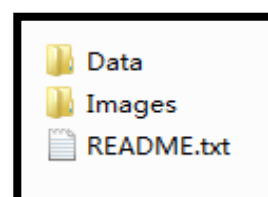


邀请了 161 位测试者对图像进行主观评分，测试图像包括 5 组(每组中含有一定数量的原始图像)，其中 JPEG 2000 文件夹 227 张图像，JPEG 文件夹 233 张，白噪声 wn 文件夹 175 张，高斯模糊 gblur 文件夹 175 张，快速瑞利衰减失真 fastfading 文件夹 175 张，共计 982 张图像。

LIVE 库中的 DMOS 值取值范围为[0, 100]，值越大图像质量越差，零值表示无失真图像。

1.2 LIVE in the Wild Image Quality Challenge database

LIVE in the Wild Image Quality Challenge database 包含 1162 张失真图像及其人眼打分值 MOS，其中的全部图像都是由现代移动设备拍摄得到的存在“真



实失真”的图像，所谓“真实失真”是指，在自然状况下经过多种非人为的图像处理过程而造成的多种混合失真并存于一幅图像的失真情况。所以不同于 LIVE，该数据库不再以失真类型作为分类依据。

在人眼打分阶段，设计了一个新型的在线众包系统，使得参与打分的实验人员由随机的网上大众构成，并最终获得了来自 8100 名参与者的共 350000 个主观分数。

LIVE 库中的 MOS 值取值范围为[0, 100]，值越大图像质量越好，满分表示无失真图像。

1.3 Metrics

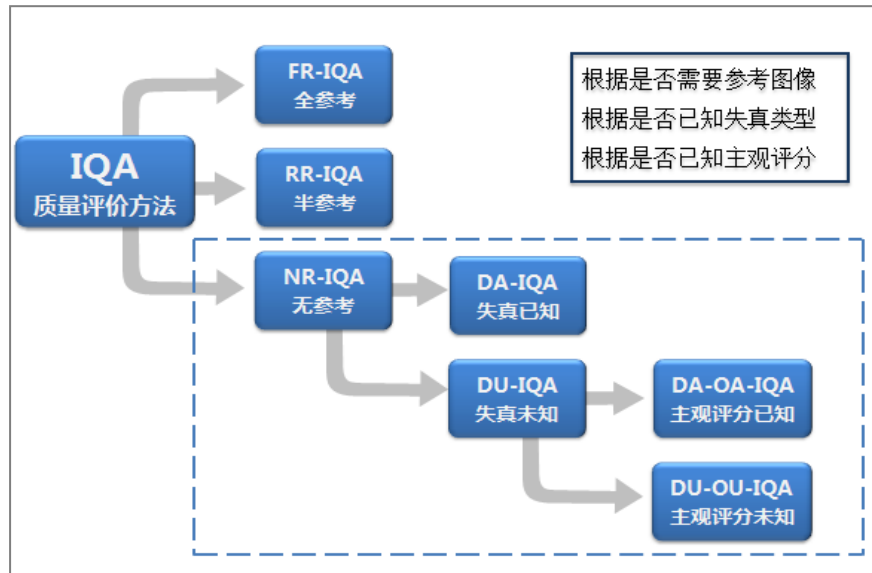
图像质量评价算法的评价效果好坏需要衡量标准，本文用到的最常用的四种为：均方根误差(RMSE) ,平均绝对误差(MAE) ,皮尔逊线性相关系数(PLCC) ,斯皮尔曼秩相关系数 (SROCC) ,

其中，PLCC 用来描述算法评价值与人眼主观打分之间的相关性，主要衡量了算法预测的准确性， SROCC 主要用来衡量算法预测的单调性。

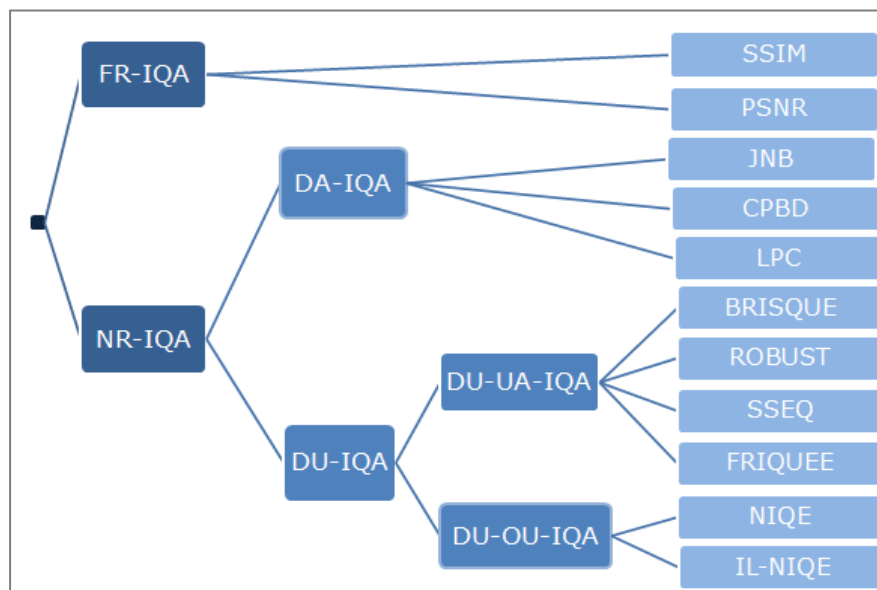
前两者越小越好，后两者越大（接近 1）越好。

2 Image Quality Assessment

现有算法为：SSIM^[1]、PSNR、JNB^[2]、CPBD^[3]、LPC^[4]、BRISQUE^[5]、ROBUST^[6]、SSEQ^[7]、FRIQUEE^[8]、NIQE^[9]、IL-NIQE^[10]。具体分类如图：



每种类别所对应的具体算法如图，接下来的实验结果分析也已此图作为分类依据：

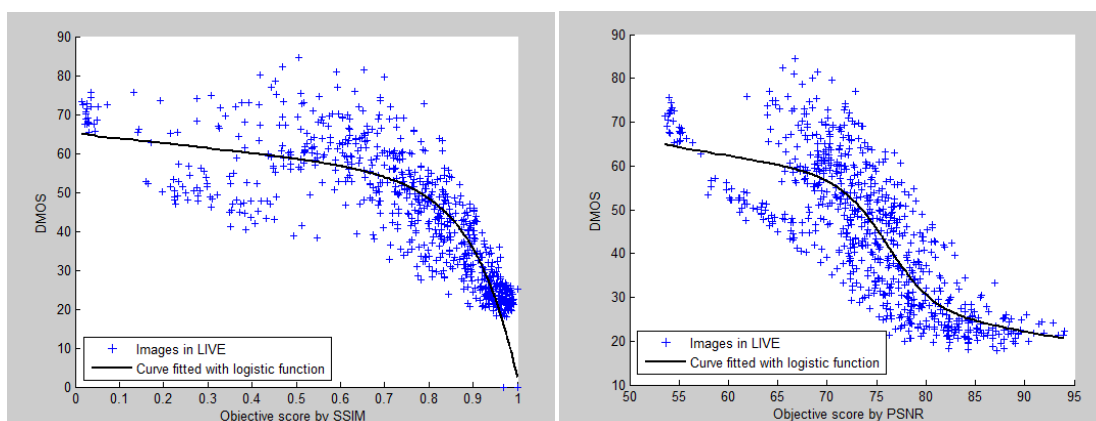


3 Experimental Results

3.1 FR-IQA (SSIM、PSNR)

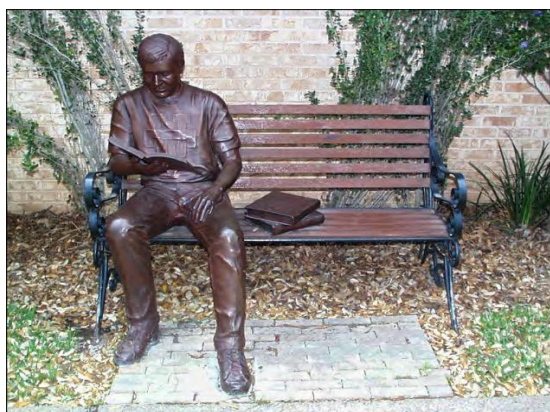
SSIM、PSNR 为全参考图像质量评价方法 (FR-IQA), 这类方法需要原始图像作为参考, 来对待测图像进行评分, 所以 LIVE 数据集必须按照 29 种图像重新进行分类, 并对每一种图像分别进行评价。结果如下:

Image	Metric	SSIM				PSNR			
		PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
1	bikes	0.9525	0.9476	6.8700	5.8176	0.8860	0.8648	7.1759	5.7523
2	building2	0.9681	0.9590	6.1131	5.0766	0.8903	0.8794	7.8026	6.1412
3	buildings	0.9730	0.9479	5.2481	4.2745	0.8822	0.8418	7.0505	5.7760
4	caps	0.9531	0.8764	6.8560	4.3715	0.8502	0.8135	7.1609	5.5084
5	carnivaldolls	0.9379	0.9184	8.2499	6.9506	0.8752	0.8337	8.2522	6.8786
6	cemetery	0.9614	0.9311	6.8319	5.2969	0.8508	0.8200	8.2778	7.0977
7	urchandcapi	0.9515	0.9278	8.1868	6.4118	0.8883	0.8680	8.8068	7.1508
8	binsinfountain	0.9526	0.9414	6.8292	5.0860	0.9316	0.8701	5.3921	3.7879
9	dancers	0.9597	0.9176	6.5596	5.0343	0.8750	0.8005	6.8592	5.2901
10	lowersonih3	0.9419	0.9557	8.2205	6.9771	0.9126	0.9164	7.0711	5.6788
11	house	0.9490	0.9289	6.9660	5.6442	0.8942	0.8605	7.0674	5.6696
12	lighthouse	0.9460	0.9517	7.4096	5.3392	0.8645	0.8492	7.9442	6.1602
13	lighthouse2	0.9400	0.9105	7.2488	5.8085	0.9163	0.8321	6.0034	4.4618
14	manfishing	0.9519	0.9474	6.9223	5.6377	0.8748	0.8596	7.6494	6.6154
15	monarch	0.9190	0.9110	8.7076	6.1079	0.8950	0.8754	6.6564	4.6886
16	ocean	0.9390	0.9649	7.7197	6.4969	0.8997	0.9274	7.2657	5.8721
17	paintedhouse	0.9321	0.9290	8.9139	6.8670	0.7940	0.7959	10.2211	8.4908
18	parrots	0.9370	0.8994	7.6992	4.8205	0.8397	0.8539	8.3503	6.4767
19	plane	0.9285	0.9184	8.3944	6.8013	0.8735	0.8304	7.9495	6.2195
20	rapids	0.9513	0.9278	7.6444	5.5162	0.8634	0.8332	8.7334	6.5139
21	sailing1	0.9535	0.9459	6.8873	5.3811	0.9011	0.8724	6.9879	5.8413
22	sailing2	0.9292	0.9187	8.0459	5.3873	0.8713	0.8651	7.4355	5.1351
23	sailing3	0.9490	0.9397	7.1083	4.9489	0.8886	0.8961	7.3537	5.6088
24	sailing4	0.9516	0.8786	6.8600	5.0137	0.8352	0.7482	7.4444	6.0976
25	statue	0.9497	0.9332	6.9556	5.5119	0.8880	0.8646	7.4938	5.7959
26	stream	0.9610	0.9428	6.7597	5.2533	0.9130	0.8619	7.4886	5.4667
27	udentsculptu	0.9623	0.9779	6.0812	5.2533	0.9246	0.9303	5.8773	4.3893
28	woman	0.9639	0.9355	6.0686	4.5169	0.9167	0.8620	6.2330	4.8603
29	womanhat	0.9329	0.9105	7.7224	6.0519	0.8639	0.8236	7.0702	5.7570
ALL		0.9371	0.9252	8.0691	6.3753	0.8076	0.8018	9.4978	7.7032



✧ 结果分析：

- (1) SSIM (Structural Similarity) 是基于结构相似度的 FR-IQA , 它认为自然图像信号是高度结构化的, 即像素之间具有很强的相关性, 这种相关性蕴含着视觉场景中物体结构的重要信息。由运行结果也知, LIVE 数据集中 ‘buildings’、 ‘student sculpture’ 这类结构信息明显的图像, 其评价结果最好。(2004)



- (2) PSNR (Peak Signal Noise Ratio) 是基于峰值信噪比的 FR-IQA , 需要注意的是, 因为 LIVE 数据库的 982 张图片中包含 203 张原图, PSNR 的结果值是 inf, 即为无穷大, 所以我计算的时候去除了这 203 个数据。
- (3) 由拟合曲线可知: 曲线与散点分布趋势大体一致, 说明算法效果还可以。
- (4) 性能比较: $SSIM > PSNR$
- (5) 速度比较: $PSNR > SSIM$

3.2 NR-DA-IQA (JNB、CPBD、LPC)

No Reference - Distortion Aware - IQA 为专用型无参考图像质量评价方法，即面向特定失真的评价方法，其中 JNB、CPBD、LPC 为三种典型的适合评价模糊失真的 IQA。将上述方法分别对 LIVE 中的 5 类失真图像类型进行评价，运行结果如下（红色标注为效果最好）：

JNB (Just Noticeable Blur 基于人眼视觉的可察觉模糊)

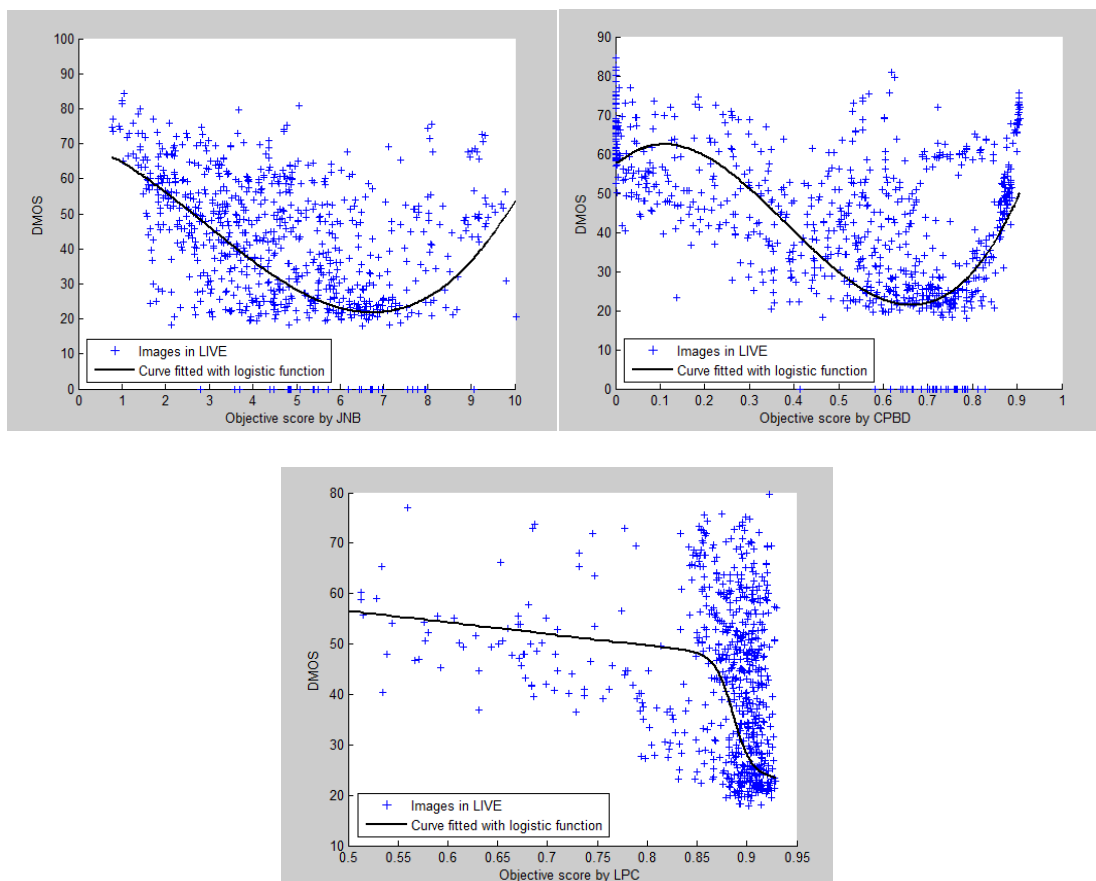
	JNB					
	jp2k	jpeg	wn	gblur	ff	ALL
PLCC	0.7153	0.4959	0.6531	0.8422	0.7424	0.4654
SROCC	0.7220	0.4123	0.6435	0.8437	0.7380	0.3996
RMSE	17.0490	21.0538	16.6455	11.7872	14.8010	20.4774
MAE	13.8162	17.4032	12.7193	9.3549	11.6047	16.5430

CPBD (Cumulative Probability of Blur Detection 模糊检测累加概率)

	CPBD					
	jp2k	jpeg	wn	gblur	ff	ALL
PLCC	0.8832	0.3251	-0.8513	0.9136	0.7944	0.5123
SROCC	0.8875	0.4029	-0.9405	0.9436	0.8125	0.3609
RMSE	11.4397	22.9270	21.9816	8.8912	13.4197	19.8690
MAE	9.0474	19.4195	18.2560	6.8564	10.0710	15.9447

LPC (Local Phase Coherence 局部相位一致性)

	LPC					
	jp2k	jpeg	wn	gblur	ff	ALL
PLCC	0.4007	0.2471	0.7329	0.9263	0.7888	0.5045
SROCC	0.3637	0.2225	0.7285	0.9613	0.8077	0.4945
RMSE	22.3511	24.2443	14.9550	8.2352	13.5791	19.9742
MAE	18.6588	21.4356	12.0456	6.7560	10.9673	16.4157



✧ 结果分析：

- (1) JNB 是基于边缘分析方法的图像模糊度评价，这个方法评价高斯模糊 gblur 失真图像效果最好。(2009)
- (2) CPBD 是在 JNB 基础上改进的，它是通过计算边缘块的边缘展宽和亮度对比度，结合 JNB，统计累加概率得到模糊度。其效果明显好于 JNB。
- (3) LPC 是基于变换域方法的图像模糊度评价，在复小波域计算局部相位一致性图，一致性越大权重越大并计算加权和，得到图像质量。(2010)
- (4) 由拟合曲线可知：这三种算法在整体的 LIVE 数据集上效果不好，因为它们都是专用型 IQA。所以他们不适合评价除高斯模糊失真以外的失真图像。
- (5) 性能比较：LPC > CPBD > JNB
- (6) 速度比较：CPBD > LPC > JNB

3.3 NR-DU-OA-IQA (BRISQUE、ROBUST、SSEQ、FRIQUEE)

No Reference - Distortion Unaware – Opinion Aware - IQA 为通用型无参考图像质量评价方法，即面向非特定失真的评价方法。

其中 BRISQUE、ROBUST、SSEQ、FRIQUEE 为 4 种基于支持向量机的 IQA，这表明这些方法都需要训练集来进行模型训练，训练集有 Opinion Aware 和 Opinion Unaware 两种。

OA 表示训练集中每张图像所对应的人眼平均打分值 (DMOS) 是已知的，所以我们可以提取图像空间域或变换域特征，基于已知质量分数的数据训练一个回归模型，然后评测未知图像的质量；

OU 表示我们不知道训练集中图像的 DMOS 值，所以在训练回归模型时，直接在原始图像上提取图像特征，放入多元高斯模型中训练得到原始 MVG 模型，然后通过评价待测图像的 MVG 模型参数与预先建立的 MVG 模型参数之间的巴氏距离来评测未知图像的质量。

注意：OA 的数据集包括原始参考图像和失真图像及其 dmos 值，而 OU 的数据集只能由一系列高清图像来组成，不能含有失真图像。

上述 4 种 OU 方法运行结果如下：

BRISQUE (Blind/ Referenceless Image Spatial Quality Evaluator

盲/无参考图像空域质量评价)

	BRISQUE					
	jp2k	jpeg	wn	gblur	ff	ALL
PLCC	0.8613	0.8218	0.8731	0.9383	0.8260	0.8255
SROCC	0.8508	0.8116	0.8951	0.9466	0.8302	0.8365
RMSE	12.3961	13.8125	10.7173	7.5612	12.4532	13.0573
MAE	10.0740	11.1237	8.0344	5.9170	9.1679	10.3478

ROBUST (Robust Statistical Approach 鲁棒性统计方法)

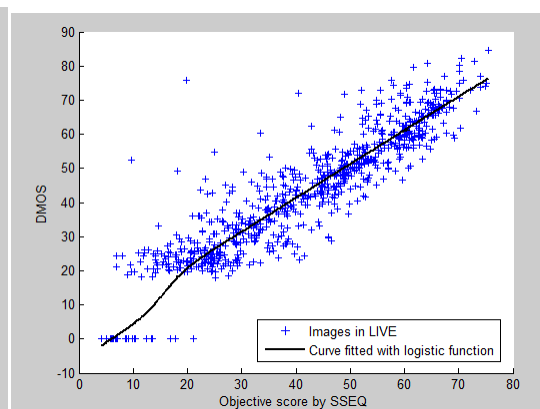
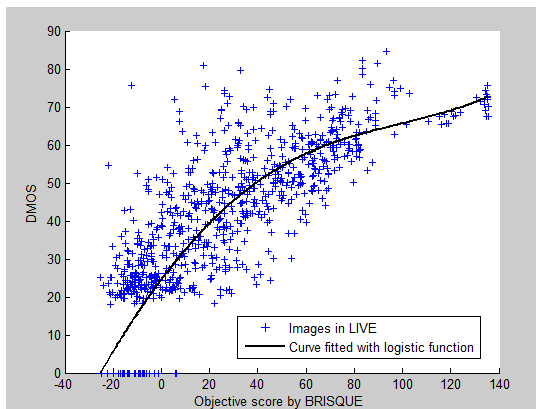
		ROBUST					
		jp2k	jpeg	wn	gblur	ff	ALL
PLCC	min	0.8299	0.8935	0.9515	0.7296	0.6372	0.8578
	max	0.9869	0.9922	0.9979	0.9935	0.9906	0.9772
SROCC	min	0.8066	0.8520	0.9288	0.6454	0.5675	0.8442
	max	0.9711	0.9864	0.9960	0.9875	0.9666	0.9736

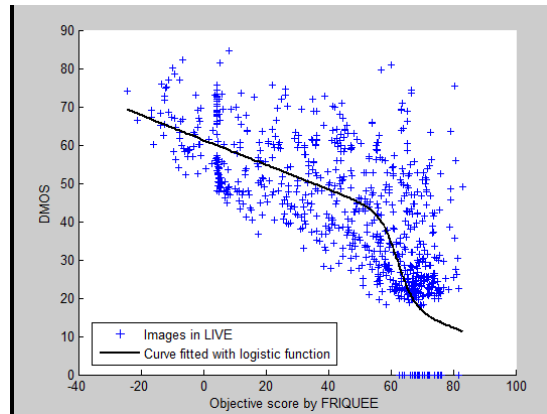
SSEQ (Spatial and Spectral Entropies 基于空间熵和谱熵)

		SSEQ					
		jp2k	jpeg	wn	gblur	ff	ALL
PLCC		0.9435	0.9715	0.9679	0.9608	0.8999	0.9421
SROCC		0.9418	0.9589	0.9817	0.9541	0.8958	0.9447
RMSE		8.0816	5.7463	5.5289	6.0609	9.6354	7.7556
MAE		6.2141	4.2204	3.9351	4.6273	6.7455	5.5128

FRIQUEE (Feature maps based Referenceless Image Quality Evaluation Engine 基于特征图谱)

		FRIQUEE					
		jp2k	jpeg	wn	gblur	ff	ALL
PLCC		0.8416	0.6233	0.9087	0.9280	0.8568	0.7496
SROCC		0.8426	0.5740	0.9717	0.9416	0.8653	0.7563
RMSE		13.1773	18.9586	9.1779	8.1440	11.3916	15.3130
MAE		10.8889	14.6041	7.9631	6.4094	9.0951	12.4069





◇ **结果分析：**

- (1) BRISQUE 是基于空域统计特征的 IQA ,评价各种失真类型的结果很平均 ,
其中评价高斯模糊 gblur 失真图像准确率稍高一些。(2011)
- (2) ROBUST 是在 BRISQUE 基础上 , 结合 L-moments 线性矩的 IQA , 这
个方法第一次将稳健性统计用于图像质量评价。评价白噪声 wn 失真效果最
好 , 且整体效果高于 BRISQUE,但运行速度较慢。(2012)
- (3) SSEQ 提取待测图像的 6 维空间熵特征和 6 维谱熵特征 进行质量评价。
是目前为止不论在不同失真类型上还是在整个 LIVE 数据集上效果都是最好
的 IQA,且运行速度也很快。(2014)
- (4) FRIQUEE 是基于特征图谱的 IQA ,该方法从一张照片中建立多个彩色空
间和变换域的特征图谱 , 然后从每个图谱中分别提取场景统计信息 , 然后进
行质量评价。**值得注意的是：** 该方法使用了 ‘LIVE in the Wild Image
Quality Challenge Database’ (2015) , 这个数据集里的 1162 张图像皆
取自智能手机和平板电脑等设备拍摄的照片 , 其失真类型及失真原因更加复
杂 , 更接近真实世界中存在的失真。(2016)
- (5) 由拟合曲线可知：这三种方法整体的效果都不错 , 散点的分布也比较均
匀 , 其中尤以 SSEQ 方法性能最好 , 几乎呈现的是线性分布。

(6) 性能比较：SSEQ>ROBUST>BRISQUE (>FRISQUEE)

(7) 速度比较：BRISQUE> SSEQ>ROBUST (>FRISQUEE)

3.4 NR-DU-OU-IQA (NIQE、IL-NIQE)

No Reference - Distortion Unaware – Opinion Unaware – IQA 为通用型无参考图像质量评价方法，且不知待测图像失真类型及训练集的人眼打分值。

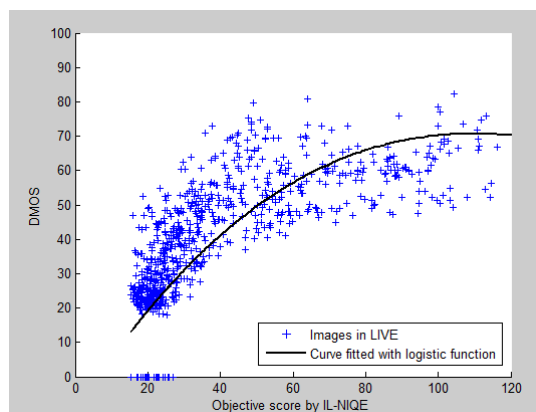
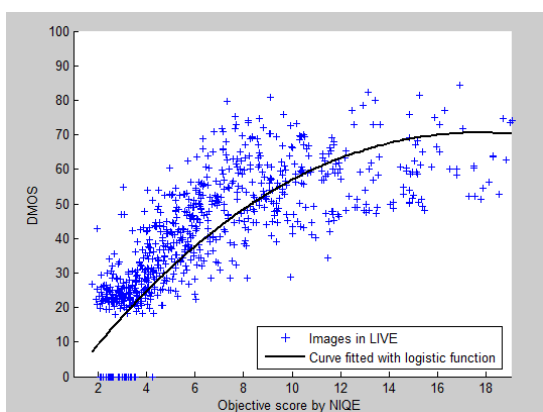
NIQE、IL-NIQE 都是在一个由一系列高清图像构成的数据集上进行训练，然后分别对 LIVE 中的 5 类失真图像类型进行评价，运行结果如下：

NIQE (Natural Image Quality Evaluator 自然图像质量评价)

	NIQE					
	jp2k	jpeg	wn	gblur	ff	ALL
PLCC	0.9101	0.8699	0.9546	0.9139	0.8722	0.8672
SROCC	0.9069	0.8436	0.9748	0.9262	0.8824	0.8724
RMSE	10.1112	11.9599	6.5507	8.8274	10.8070	11.5102
MAE	8.4251	9.8231	5.2050	6.3473	8.2780	9.5104

IL-NIQE (Integrated Local NIQE 局部综合的自然图像质量评价)

	IL-NIQE					
	jp2k	jpeg	wn	gblur	ff	ALL
PLCC	0.8606	0.8677	0.9340	0.8651	0.7951	0.8364
SROCC	0.8542	0.8405	0.9716	0.8980	0.7872	0.8446
RMSE	12.4239	12.0511	7.8549	10.9650	13.3990	12.6794
MAE	10.4791	9.9675	6.5840	8.6271	10.3270	10.5827



✧ 结果分析：

- (1) NIQE 是基于概率模型的 IQA , 评价 LIVE 各类失真效果都不错 , 其中以白噪声 wn 效果最好 , jpeg 较差。(2012)
 - (2) IL-NIQE 是在 NIQE 基础上的改进方法 , 增加了 3 种 NSS 特征 , 且将待测图像划分为图像子块 , 将每个子块都用一个 MVG 模型描述 , 衡量与预先建立的 MVG 模型参数之间的距离并计算每个图像块的质量 , 然后进行局部综合。由运行结果来看 , 对于白噪声 wn 失真评价效果最好 , jp2k 较差。
 - (3) 由拟合曲线可知 : 这两种方法的散点分布较为均匀 , 其曲线趋势也几乎一致 , 说明它们在整个 LIVE 集上的效果也都不错。
 - (4) 性能比较 : NIQE>IL-NIQE
 - (5) 速度比较 : NIQE>IL-NIQE
- (这里出现与论文不符合的情况 , IL-NIQE 本应效果更好 , 但实际上效果却不如 NIQE。)

4 Conclusions

将上述 11 种方法在 LIVE 上的评价结果汇总 ,只考虑两个线性系数指标 PLCC 和 SROCC ,结果如图 (红色标注为效果最好的算法 , 橘红色标注为效果第二的算法):

PLCC (Pearson linear correlation coefficient 皮尔逊线性相关系数)

		PLCC					
		jp2k	jpeg	wn	gblur	ff	ALL
FR	SSIM	-	-	-	-	-	0.9371
	PSNR	-	-	-	-	-	0.8076
NR	JNB	0.7153	0.4959	0.6531	0.8422	0.7424	0.4654
	CPBD	0.8832	0.3251	-0.8513	0.9136	0.7944	0.5123
	LPC	0.4007	0.2471	0.7329	0.9263	0.7888	0.5045
	BRISQUE	0.8613	0.8218	0.8731	0.9383	0.8260	0.8255
	SSEQ	0.9435	0.9715	0.9679	0.9608	0.8999	0.9421
	FRIQUEE	0.8416	0.6233	0.9087	0.9280	0.8568	0.7496
	NIQE	0.9101	0.8699	0.9546	0.9139	0.8722	0.8672
	IL-NIQE	0.8606	0.8677	0.9340	0.8651	0.7951	0.8364

SROCC (Spearman rank ordered correlation coefficient

斯皮尔曼秩相关系数)

		SROCC					
		jp2k	jpeg	wn	gblur	ff	ALL
FR	SSIM	-	-	-	-	-	0.9252
	PSNR	-	-	-	-	-	0.8018
NR	JNB	0.7220	0.4123	0.6435	0.8437	0.7380	0.3996
	CPBD	0.8875	0.4029	-0.9405	0.9436	0.8125	0.3609
	LPC	0.3637	0.2225	0.7285	0.9613	0.8077	0.4945
	BRISQUE	0.8508	0.8116	0.8951	0.9466	0.8302	0.8365
	SSEQ	0.9418	0.9589	0.9817	0.9541	0.8958	0.9447
	FRIQUEE	0.8426	0.5740	0.9717	0.9416	0.8653	0.7563
	NIQE	0.9069	0.8436	0.9748	0.9262	0.8824	0.8724
	IL-NIQE	0.8542	0.8405	0.9716	0.8980	0.7872	0.8446

◇ **综上：**

- (1) 不同算法的性能之间有较大差异。就上述结果来看，性能最好的质量评价算法是 SSEQ，其次是 NIQE，它们的预测精准度（PLCC）最高，而且其预测的一致性（SROCC）最好。
- (2) 训练集的不同可能会影响测试结果。本文提到的算法中，需要支持向量机进行训练的算法有 6 种：BRISQUE、ROBUST、SSEQ 的训练集是 LIVE database；FRIQUEE 的训练集是 LIVE in the Wild Image Quality Challenge Database；NIQE、ILNIQE 的训练集是算法自带的一个高清图片数据集。所以在进行算法之间的比较时需要考虑到训练集的数据构成。
- (3) 不同算法的计算复杂度（平均处理时间）相差较大。本文提到的 11 种算法中，质量评价速度最快的 PSNR，其每张图片的处理时间仅需 0.12 秒；但速度最慢的 FRIQUEE，其处理时间约为 53.67 秒。运行速度可能取决于算法所提取的图像特征或其他因素。运行速度的快慢也是我们在评价算法时需要考虑到的问题。

REFERENCE

- [1] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing , vol.13, no.4, April 2004.
- [2] Rony Ferzli and Lina J. Karam," A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur(JNB)" , IEEE Tran. Image Processing, vol. 18, no. 4, April 2009.
- [3] Niranjana D. Narvekar and Lina J. Karam , " A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection" ,IEEE Trans. Image Processing, vol.20, no.9,September 2011.
- [4] Rania Hassen, Zhou Wang," Image Sharpness Assessment Based on Local Phase Coherence" , IEEE Tran. Image Processing, vol. 22, no. 7, July 2013.
- [5] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," IEEE Transactions on Image Processing , 2012.
- [6] A. Mittal, A. K. Moorthy and A. C. Bovik, " Making image quality assessment robust " Forty-Sixth Annual Asilomar Conference on Signals, Systems, and Computers, Monterey, California, November 04-07, 2012.
- [7] L. Liu, B. Liu, H. Huang, and A.C. Bovik , " No-reference image quality assessment based on spatial and spectral entropies," Signal Processing: Image Communication, June 2014.
- [8] D. Ghadiyaram and A. C. Bovik, "Perceptual Quality Prediction on Authentically Distorted Images Using a Bag of Features Approach," <http://arxiv.org/abs/1609.04757> (under review).
- [9] A. Mittal, R. Soundararajan and A. C. Bovik, " Making a Completely Blind Image Quality Analyzer, " IEEE Signal Processing Letters , pp. 209-212, vol. 22, no. 3, March 2013.
- [10] Lin Zhang, Lei Zhang, and Alan C. Bovik, "A feature-enriched completely blind image quality evaluator", IEEE Trans. Image Processing, vol. 24, no. 8, pp. 2579-2591, 2015.