

文章编号: 1001-3806(2016)06-0902-05

基于结构相似度评价方法的窗口效应

王强, 张合新*, 孟飞, 张腾飞

(第二炮兵工程大学 控制工程系, 西安 710025)

摘要: 为了提高基于结构相似度图像评价方法的评价效果,对传统的窗口选取准则进行了改进,采用实验对比的方法,通过对窗口改进前后评价效果的对比验证,得到了不同数据库和不同失真类型的最佳窗口。结果表明,传统的单一窗口并不能得到最佳的评价效果,而且现在普遍使用的标准差为 1.5 的高斯加权窗口的效果也不一定好于普通的方形窗口,不同的失真类型的图像其最优窗口也是不同的。结果表明,对于亮度窗口 W_1 来说,窗口越大效果越好;而对于对比度、结构度窗口 W_2 而言, $W_2 = 7$ 时效果最好。该研究成果对优化基于结构相似度评价方法的效果具有重要作用。

关键词: 图像处理;窗口效应;图像质量评价;结构相似

中图分类号: TN911.73

文献标志码: A

doi: 10.7510/jgjs.issn.1001-3806.2016.06.026

Window effect of evaluation method of the structural similarity

WANG Qiang, ZHANG Hexin, MENG Fei, ZHANG Tengfei

(Department of Automation, The Second Artillery Engineering University, Xi'an 710025, China)

Abstract: In order to improve evaluation effect of structural similarity index image, the traditional window selection criterion was improved. By contrasting the window evaluation value before and after improvement, the best window for different databases and different types of distortion was obtained. The results show that the traditional single window does not get the best evaluation result, and the widely-used Gaussian weighted window with standard deviation of 1.5 does not better than the ordinary square window. The optimal windows of different types of image distortion are also different. From the experimental results, for luminance window W_1 , the bigger the window, the better the evaluation result. And for contrast and structure window W_2 , the evaluation result is best when $W_2 = 7$. The results play an important role on optimizing the evaluation method of structural similarity index.

Key words: image processing; window effect; image quality assessment; structural similarity

引言

数字图像在获取、处理、压缩、存储和传输过程中会引入各种失真,这些失真会对图像的质量产生很大影响,最后直接影响到图像的观察和使用。图像质量评价的目的在于:一方面是为了分析图像在引入这些失真后对其本身影响;另一方面可以为成像系统参量设计和算法优化提供参考。图像评价分为主观和客观评价,客观评价方法因为其良好的应用性成为图像质量评价的热点。客观评价方法根据能够得到的参考信息的多少通常分为三大类,分别为全参考质量评价、半参考质量评价和无参考质量评价。无参考评价一般针

对特定的失真类型进行研究。半参考评价通常通过一个辅助通道获取参考图像部分参考信息,可靠性介于另两种评价方法之间。全参考评价可以得到参考图像的全部信息,通过对比两幅图像来进行,它是目前为止研究最多,结果最可靠的评价方法。在全参考评价方法中基于结构相似度的图像质量评价(structural similarity index method, SSIM)及其各种改进方法是使用较多、效果比较理想的一种方法。在应用这种方法进行图像质量评价的时候,早期大都使用的是 8×8 的方形窗口,后来为了防止图像产生分块效应,窗口就选用 11×11 、标准差为 1.5 的圆形对称的高斯加权窗口。至此,高斯加权窗口就一直作为 SSIM 的默认窗口。

本文中传统的 SSIM 评价参量的窗口选取规则进行了改进,并对改进后的评价方法进行了实验验证,实验结果表明,现在普遍使用的标准差为 1.5 的圆形对称的高斯加权窗口针对不同数据库、不同失真类型图像并不一定是最优的窗口。

基金项目:国家自然科学基金资助项目(61203189)

作者简介:王强(1993-),男,硕士研究生,研究方向为控制科学与工程。

* 通讯联系人。E-mail: 17792616386@163.com

收稿日期:2015-11-09;收到修改稿日期:2015-11-24

1 基于结构相似度的图像质量评价(SSIM)

自然场景具有很强的结构性, WANG 等人假设人类视觉具有提取场景结构信息的高度适应性, 并根据图像像素间的相关性, 提出了结构信息的概念, 认为结构信息独立于图像的亮度和对比度信息。并根据结构相似度的定义, 提出了基于结构相似度的图像质量评价方法^[1]。

假设 x, y 分别代表参考图像和失真图像。 x, y 的均值和标准差分别记为 μ_x, μ_y 和 σ_x, σ_y , x 和 y 的相关系数记为 σ_{xy} , ω_i 是不同像素点的权重。均值、标准差、相关系数的计算公式如下:

$$\mu_x = \sum_{i=1}^N \omega_i x_i \quad (1)$$

$$\sigma_x = \left[\sum_{i=1}^N \omega_i (x_i - \mu_x)^2 \right]^{\frac{1}{2}} \quad (2)$$

$$\sigma_{xy} = \sum_{i=1}^N \omega_i (x_i - \mu_x)(y_i - \mu_y) \quad (3)$$

式中, x 和 y 的亮度、对比度和结构相似度分别定义为:

$$l(x, y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (4)$$

$$c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (5)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (6)$$

式中, C_1, C_2 和 C_3 是为了避免 3 个结构因子的分母为零而产生奇异性所引入的 3 个很小的正数。结合这 3 种相似性准则而构成的质量评价参量定义如下^[2]:

$$S_{SSIM} = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (7)$$

当取 $C_1 = (K_1 L), C_2 = (K_2 L), C_3 = C_2/2$ (L 是图像动态范围), 系数 $K_1 \ll 1, K_2 \ll 1, \alpha = \beta = \gamma = 1$, (7) 式可以简化为:

$$S_{SSIM} = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

根据(8)式, 并通过一个滑动的窗口遍历该图像, 然后计算各个以该点为中心的图像块的 SSIM 值, 获得整幅图像关于 SSIM 的索引图^[1]。最后通过简单地对各点的值求平均, 得出对失真图像最终的评价结果。公式如下所示:

$$\bar{M}_{SSIM} = \sum_{i=1}^M \frac{S_{SSIM}(x_i, y_i)}{M} \quad (9)$$

式中, M 是窗口个数。

2 实验方法

本文将通过实验验证窗口调整对图像评价效果

的影响, 实验主要分 3 步: (1) 对传统的 SSIM 图像评价指标进行适当的改进调整; (2) 用调整后的图像评价指标在一个或者多个图像数据库中进行实验测量; (3) 统计评价结果, 分析不同窗口的图像评价方法的效果。

2.1 窗口的调整

在评估窗口改变对 SSIM 评价效果的影响时, 窗口做出了 3 种调整。

(1) 窗口区分: 标准的 SSIM 对其 3 个组成因子使用单一的窗口进行计算。本文将三者进行了区分, 亮度因子窗口 W_1 和对比度、结构因子窗口 W_2 将使用不同窗口, 最终公式成了如下形式:

$$\mu_{x,1} = \sum_{i=1}^N \omega_{1,i} x_i \quad (10)$$

$$\mu_{x,2} = \sum_{i=1}^N \omega_{2,i} x_i \quad (11)$$

$$\sigma_{x,1} = \sqrt{\sum_{i=1}^N \omega_{1,i} (x_i - \mu_{x,1})^2} \quad (12)$$

$$\sigma_{x,2} = \sqrt{\sum_{i=1}^N \omega_{2,i} (x_i - \mu_{x,2})^2} \quad (13)$$

$$\sigma_{xy}' = \sum_{i=1}^N \omega_{2,i} (x_i - \mu_{x,2})(y_i - \mu_{y,2}) \quad (14)$$

式中, N 是像素点的个数。

(8) 式中的 SSIM 变成了如下形式:

$$S_{SSIM}(x, y) = \frac{2\mu_{x,1}\mu_{y,1} + C_1}{\mu_{x,1}^2 + \mu_{y,1}^2 + C_1} \times \frac{2\sigma_{xy}' + C_2}{\sigma_{x,2}^2 + \sigma_{y,2}^2 + C_2} \quad (15)$$

(2) 窗口形状: 在试验中选用了两种窗口形状, 一种是在现在普遍使用的圆形高斯窗口, 另一种是在早期使用过的方形窗口。

(3) 窗口尺寸: 为了形成对比, 选取了一个常见的 21×21 的窗口作为对比对象。 21×21 的窗口输出结果与主观评价结果一致性非常高, 因为它能去除边缘效应对输出值的影响。在评估不同尺寸的方形窗口时候, 先固定其中一个窗口不变, 然后不断调整另一个窗口大小。对于方形窗口, 窗口的尺寸从 1 ~ 21 每次递增 2 个像素; 对于高斯窗口, 标准差从 0.5 ~ 5.5 每次递增 0.5 个像素。

2.2 IQA 数据库

全参考图像评价 (image quality assessment, IQA) 数据库包含了大量的失真图像以及与它们相应的“完美品质”的参考图像。而且数据库中每张失真图像都有一个通过心理视觉测试的主观质量评分^[1-9]。

(1) LIVE 数据库由 Texas 大学开发, 包含了 779

张失真图片;(2)A57数据库由 Cornell 大学开发,包含了 54 张失真图片;(3)Toyama 数据库由 Toyama 大学开发,包含了 168 张失真图片;(4)IVC 数据库由 L'Universit'e de Nantes 开发,包含了 185 张失真图片;(5)CSIQ 数据库由 Oklahoma State University 开发,包含 866 张失真图片;(6)TID2008 数据库由芬兰、意大利和乌克兰联合开发,包含 1700 张失真图片。

2.3 评估指标

本文中 will 使用 2.2 节中提到的图像数据库对窗口调整后的 SSIM 算法进行评估,评估的主要参量如下: Pearson 线性相关系数(Pearson correlation coefficient, PLCC)、平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE), Spearman 等级相关系数(Spearman's correlation coefficient for ranked data, SRCC)和 Kendall 等级相关系数(Kendall's correlation coefficient for ranked data, KRCC)^[10-12]。这些参量也常被用于评估 SSIM 算法。其中 PLCC, MAE 和 RMSE 指标通过客观评分和主观数据库测试分数之间的非线性映射后计算。而 SRCC 和 KRCC 都是非秩相关参量,不需要非线性映射。在算法评估时, PLCC, SRCC, KRCC 的值越大,算法的性能越好,相反, MAE 和 RMSE 则是值越小越好^[5]。各个参量具体计算如下(其中, r_i 是客观评价值, o_i 是主观评价值, $q(r)$ 是客观评价值通过非线性映射函数映射到主观评价的值, \bar{q} 和 \bar{o} 是评价值的均值)。

(1) 非线性映射函数

$$q(r) = a_1 \left\{ \frac{1}{2} - \frac{1}{1 + \exp[a_2(r_i - a_3)]} \right\} + a_4 r_i + a_5 \quad (16)$$

式中, a_1, a_2, a_3, a_4, a_5 是待测系数, 它们通过计算主、客观得分之间最大非线性相关性得到。

(2) Pearson 线性相关系数(PLCC):

$$P_{PLCC} = \frac{\sum_{i=1}^N [q(r) - \bar{q}] \times (o_i - \bar{o})}{\sqrt{\sum_{i=1}^N [q(r) - \bar{q}]^2 \times \sum_{i=1}^N (o_i - \bar{o})^2}} \quad (17)$$

PLCC 是一个度量两个变量相关程度的统计参量, 当 PLCC 取 1 时, 表示两者相关程度最高。

(3) 平均绝对误差(MAE):

$$M_{MAE} = \frac{1}{N} \sum_{i=1}^N |q(r) - o_i| \quad (18)$$

MAE 是用来度量客观评价值映射到主观评价后两者的紧密程度, 作为一个误差项, MAE 越小越好。

(4) 均方根误差(RMSE):

$$R_{RMSE} = \sqrt{\frac{1}{N} [q(r) - o_i]^2} \quad (19)$$

RMSE 与 MAE 相似, 是一个度量主观得分与客观映射值误差大小的统计值, 跟 MAE 一样, RMSE 值也是越小越好。

(5) Spearman 等级相关系数(SRCC):

$$C_{SRCC} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \times \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (20)$$

SRCC 是一个非统计测量参量, 用于估计两个变量之间的单调性。SRCC 的特点是受异常值的影响较小, 当取到 1 时表示完全等级相关。

(6) Kendall 等级相关系数(KRCC):

$$K_{KRCC} = \frac{2(N_c - N_d)}{N(N-1)} \quad (21)$$

式中, Kendall 等级相关系数是一个用来测量两个随机变量相关性的统计值; N_c 表示空间中拥有一致性的元素对数; N_d 表示空间中拥有不一致性的元素对数。它和 SRCC 一样是一个非统计测量参量, 取 1 时表示完全等级相关。

3 实验结果

根据实验方案对 6 个图像数据库进行了实验验证, 图像质量评价方法主要选用 SSIM、多尺度 SSIM (multiscale SSIM, MS-SSIM)^[13-14]、优化方形窗口 SSIM (window optimized SSIM square window, WO-SSIM-S)、优化高斯窗口 SSIM (window optimized SSIM Gaussian window, WO-SSIM-G) 这 4 种方法。

表 1 中列的是每个数据库的实验结果。根据测试结果对这 4 种方法从好到差进行了排序。从排序结果来看, 3 种改进的 SSIM 评价结果都要好于传统的方法。从表中还能看出, 各个数据库对于最佳的方形窗口尺寸和最佳高斯窗口的参量不是完全相同的, 这说明不同的数据库采用不同的窗口评价的效果也不同, 而且现在普遍使用的圆形高斯加权窗口评价效果不一定优于方形窗口。

表 2 中所列的结果是对所有数据库测试的结果。对于 WO-SSIM-S 和 WO-SSIM-G 列的是最佳窗口的测试结果。从这两种评价方法的最佳窗口来看, 方形亮度窗口 $W_1 = 21$, 对比度、结构窗口 $W_2 = 7$ 时的结果要好于 3 个因子采用同一个 21×21 的方形窗口时的结果。对于圆形高斯窗口分别采用标准差为 5.5 和 1.5 的窗口要好于全部使用标准差为 1.5 的窗口。

表 3 中根据不同的失真类型对 LIVE, CSIQ 进行了测试, 并列出了不同失真类型的优化窗口。从结果中可以很明显看出, 不同的数据库不同的失真类型最

Table 1 Evaluation results of six databases

LIVE database: $S(1,9) G[0.5, 2.0]$					
metric	PLCC	MAE	RMS	SRCC	KRCC
MS-SSIM	0.9489	6.6980	8.6190	0.9513	0.8044
WO-SSIM-S	0.9477	6.7490	8.7224	0.9504	0.8021
WO-SSIM-G	0.9460	6.8681	8.8573	0.9486	0.7981
SSIM	0.9449	6.9330	8.9460	0.9479	0.7963
A57 database: $S(21,21) G[4.0, 5.5]$					
metric	PLCC	MAE	RMS	SRCC	KRCC
WO-SSIM-S	0.8750	0.0940	0.1190	0.8725	0.6958
WO-SSIM-G	0.8645	0.0969	0.1236	0.8649	0.6860
MS-SSIM	0.8603	0.1007	0.1253	0.8414	0.6478
SSIM	0.8017	0.1209	0.1469	0.8066	0.6058
TOYAMA database: $S(7,9) G[1.0, 2.5]$					
metric	PLCC	MAE	RMS	SRCC	KRCC
WO-SSIM-S	0.9192	0.4239	0.5549	0.8912	0.7353
WO-SSIM-G	0.9149	0.4300	0.5634	0.8883	0.7286
MS-SSIM	0.9119	0.4328	0.5640	0.8874	0.7223
SSIM	0.9108	0.4386	0.5738	0.8794	0.7203
IVC database: $S(21,7) G[2.0, 5.5]$					
metric	PLCC	MAE	RMS	SRCC	KRCC
MS-SSIM	0.9192	0.3666	0.4797	0.9091	0.7353
WO-SSIM-S	0.9149	0.3727	0.4918	0.9042	0.7286
WO-SSIM-G	0.9119	0.3777	0.4999	0.9018	0.7223
SSIM	0.9108	0.3813	0.5029	0.8980	0.7203
CSIQ database: $S(21,5) G[5.5, 1.5]$					
metric	PLCC	MAE	RMS	SRCC	KRCC
MS-SSIM	0.8991	0.0870	0.1149	0.9133	0.7393
WO-SSIM-S	0.8887	0.0900	0.1204	0.9032	0.7219
WO-SSIM-G	0.8808	0.0933	0.1243	0.8950	0.7126
SSIM	0.8612	0.0992	0.1334	0.8756	0.6907
TID2008 database: $S(21,7) G[5.5, 1.5]$					
metric	PLCC	MAE	RMS	SRCC	KRCC
MS-SSIM	0.8451	0.5578	0.7173	0.8542	0.6568
WO-SSIM-S	0.7949	0.6330	0.8146	0.7973	0.5965
WO-SSIM-G	0.7888	0.6391	0.8248	0.7934	0.5926
SSIM	0.7732	0.6546	0.8511	0.7749	0.5768

Table 2 Weighted averages of PLCC, SRCC, and KRCC of six databases

all databases: $S(21,7) G[5.5, 1.5]$			
metric	PLCC	SRCC	KRCC
MS-SSIM	0.8847	0.8914	0.7116
WO-SSIM-S	0.8581	0.8620	0.6786
WO-SSIM-G	0.8533	0.8586	0.6742
SSIM	0.8416	0.8455	0.6615

佳评价窗口都有所不同。

图 1 和图 2 表现了不同窗口对评价结果的影响。图中只画了方形窗口优化的对结果的影响,高斯窗口优化结果与此相似。画图时使用来自所有数据库的加权平均 KRCC 结果生成上述图像。可以发现两幅图

Table 3 Optimal window sizes for distortion images of LIVE and CSIQ databases

distortion	LIVE	CSIQ
JPEG	(3,9)	(21,9)
JPEG2000	(3,7)	(5,15)
blur	(3,7)	(3,17)
noise	(1,3)	(21,3)

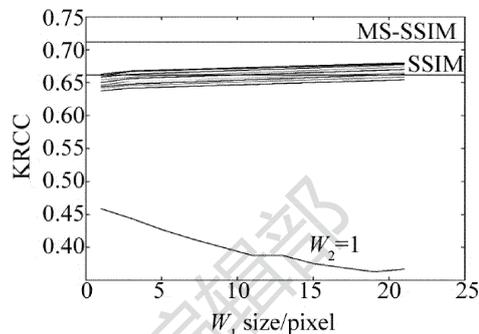


Fig. 1 Relationship between KRCC and W_1 when W_2 is fixed

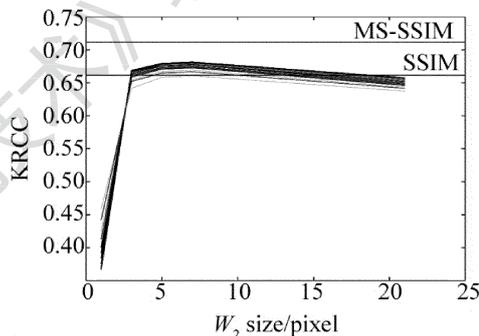


Fig. 2 Relationship between KRCC and W_2 when W_1 is fixed

像的曲线总体走势是不一样的。在图 1 中随着 W_1 (亮度窗口) 大小的增加曲线是不断增加的。而图 2 中在曲线的峰值之后,随着 W_2 (对比度、结构窗口) 的增大曲线是在缓慢下降的。从图中可以总结得到在实验的取值范围内,除了 $W_2 = 1, W_1$ 越大评价效果越好;而对于 W_2 来讲, $W_2 = 7$ 时评价效果则是最佳的。

在表 4 中对 3 种图像评价方法的性能进行了对比。从每张图像评价用时看,WO-SSIM 的计算比 SSIM 复杂 20%,这是由于计算两个单独的 μ 值造成的。所以说提高评价的效果是以牺牲评价的效率为代价的^[15]。

Table 4 Performance comparison of SSIM, window optimized SSIM and multi-scale SSIM

metric	SSIM	WO-SSIM	MS-SSIM
seconds/image	0.0741	0.0890	0.2873
ratio	1.0000	1.2006	3.8778

4 结 论

从窗口的选择对 SSIM 方法评价结果的影响出发,对 SSIM 方法的窗口进行了适当的调整,并用 6 个

图像数据库进行了实验验证。从实验结果来看,单一的窗口并不能得到最佳的评价效果,而且现在普遍使用的标准差为 1.5 的高斯加权窗口的效果也不一定好于普通的方形窗口。在针对特定的失真类型的图像进行实验时发现,不同的失真类型的图像其最优窗口也是不同的。总的来说,对于亮度窗口,窗口越大,效果越好;而对于对比度、结构度窗口来说, $W_2 = 7$ 时效果最好。最后分析了 3 种不同的评价方法,从结果来看,评价效果提升的同时增加了计算复杂度,也就是说会牺牲图像评价的效率。

参 考 文 献

- [1] WANG Z, BOVIK A, SHEIKH H, *et al.* Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 4(13): 600-612.
- [2] WANG Z, BOVIK A. A universal image quality index[J]. IEEE Signal Processing Letters, 2002, 3(9): 81-84.
- [3] MIAO Y, YI S L, HE J F. Image quality assessment of feature similarity combined with gradient information[J]. Journal of Image and Graphics, 2015, 20(6): 749-755 (in Chinese).
- [4] SHEIKH H, SABIR M, BOVIK A. A statistical evaluation of recent full reference image quality assessment algorithms[J]. IEEE Transactions on Image Processing, 2006, 11(15): 3440-3451.
- [5] ZHANG Z D, CHEN J, WANG G W, *et al.* Evaluation method of super-resolution restoration based on SSIM NCCDFT [J]. Chinese Journal of Liquid Crystals and Displays, 2015, 30(4): 713-721 (in Chinese).
- [6] SHEIKH H, WANG Z, CORMACK L, *et al.* LIVE image quality assessment database release [EB/OL]. (2002-12-14) [2004-01-12]. <http://live.ece.utexas.edu/research/quality>.
- [7] YUAN Y C, LIU Y P, GAO H W. Image quality assessment method based on edge structure similarity [J]. Application Research of Computers, 2015, 32(9): 2870-2873 (in Chinese).
- [8] CHANDLE D, HEMAMI S. VSNR: A wavelet-based visual signal-to-noise ratio for natural images [EB/OL]. (2005-12-14) [2006-01-12]. <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.
- [9] SAZZAD Z, KAWAYAKE Y, HORITA Y. Image quality evaluation database [EB/OL]. (2009-05-16) [2009-12-22]. [http://mict.eng.u-toyama.ac.jp/database to yama](http://mict.eng.u-toyama.ac.jp/database%20to%20yama).
- [10] le CALLET P, AUTRUSSEAU F. Subjective quality assessment IRCYCYN/IVC database [EB/OL]. (2004-08-10) [2006-05-02]. <http://www.ircycyn.ec-nantes.fr/ivc/>.
- [11] SHI M D, YAO J H, BAI T C. Algorithm of image quality evaluation with gabor transfer function similarity measure [J]. Bulletin of Science and Technology, 2014, 30(10): 133-135 (in Chinese).
- [12] JIANG W F, YANG P Z, ZHANG H Y, *et al.* Image quality evaluation based on human visual [J]. Computer Simulation, 2015, 32(5): 213-217 (in Chinese).
- [13] LARSON E C, CHANDLER D M. Most apparent distortion: full-reference image quality assessment and the role of strategy [J]. Journal of Electronic Imaging, 2010, 19(1): 6-11.
- [14] PONOMARENKO N, LUKIN V, ZEINSKY A, *et al.* TID2008-a database for evaluation of full-reference visual quality assessment metrics [J]. Advances of Modern Radioelectronics, 2009, 10(5): 30-45.
- [15] LI Q, LIU Z, NAN B B. Improved image super-resolution reconstruction based neighbor embedding [J]. Laser Technology, 2015, 39(1): 13-18 (in Chinese).