

Accepted Manuscript

A critical survey of state-of-the-art image inpainting quality assessment metrics

Muhammad Ali Qureshi, Mohamed Deriche, Azeddine Beghdadi, Asjad Amin

PII: S1047-3203(17)30180-3

DOI: <http://dx.doi.org/10.1016/j.jvcir.2017.09.006>

Reference: YJVCI 2055

To appear in: *J. Vis. Commun. Image R.*

Received Date: 1 January 2017

Revised Date: 2 August 2017

Accepted Date: 8 September 2017



Please cite this article as: M.A. Qureshi, M. Deriche, A. Beghdadi, A. Amin, A critical survey of state-of-the-art image inpainting quality assessment metrics, *J. Vis. Commun. Image R.* (2017), doi: <http://dx.doi.org/10.1016/j.jvcir.2017.09.006>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A critical survey of state-of-the-art image inpainting quality assessment metrics

Muhammad Ali Qureshi^{a,b,*}, Mohamed Deriche^a, Azeddine Beghdadi^c, Asjad Amin^{a,b}

^a*King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia*

^b*The Islamia University of Bahawalpur, Pakistan*

^c*L2TI, Institut Galilée, Université Paris 13, Sorbonne Paris Cité, France*

Abstract

Image inpainting is the process of restoring missing pixels in digital images in a plausible way. Research in image inpainting has received considerable attention in different areas, including restoration of old and damaged documents, removal of undesirable objects, computational photography, retouching applications, etc. The challenge is that the recovery processes themselves introduce noticeable artifacts within and around the restored image regions. As an alternative to subjective evaluation by humans, a number of approaches have been introduced to quantify inpainting processes objectively. Unfortunately, existing objective metrics have their own strengths and weaknesses as they use different criteria. This paper provides a thorough insight into existing metrics related to image inpainting quality assessment, developed during the last few years. The paper provides, under a new framework, a comprehensive description of existing metrics, their strengths, their weaknesses, and a detailed performance analysis on real images from public image inpainting database. The paper also outlines future research directions and applications of inpainting and inpainting-related quality assessment measures.

Keywords: Image inpainting, Image quality assessment, Inpainting quality, Inpainting databases, Image inpainting quality assessment, Survey

*Corresponding author:

Email address: ali.qureshi@iub.edu.pk (Muhammad Ali Qureshi)

10 **1. Introduction**

11 Image inpainting is generally defined as the process of restoring missing pixels and damaged
12 regions, or removing unwanted objects in digital images in a plausible way [1]. Considerable re-
13 search has been carried in developing inpainting algorithms, and a plethora of image inpainting
14 algorithms have been proposed [1–4]. Image inpainting has recently received considerable atten-
15 tion in different areas related to image processing. While the applications of image inpainting are
16 countless, we outline below the most common and practical ones.

- 17 • *Removing Unwanted Objects:* Unwanted objects can be removed from the image using
18 inpainting techniques. The scenario is seen as a special class of image tampering. Fig. 1
19 shows a nice example of image inpainting where the cage in the original image is removed
20 in the inpainted image [5].
- 21 • *Restoring Photos:* The deterioration in photos with the passage of time can be overcome
22 using inpainting. The scratches in the photos resulting from improper handling can also be
23 removed. This is also the case of restoring images from cultural archives, etc. Fig. 2 shows
24 an example in which the scratches in the old photograph have been removed using inpainting
25 [1].
- 26 • *Photo Retouching:* Another widely used application of image inpainting is in the media
27 industry where photos of actors/actresses, models, etc., are manipulated by removing wrin-
28 kles, mole marks, or undesirable facial features to make these “more attractive”. Fig. 3
29 shows an example of image inpainting where the face is made more attractive by removing
30 some marks using inpainting [6].
- 31 • *Text Removal:* Image inpainting can also be used for removing unwanted text, stamps,
32 copyright logos, etc., in digital images. Fig. 4 shows an example of a street image with
33 superimposed text, from which the text is removed in the inpainted image.

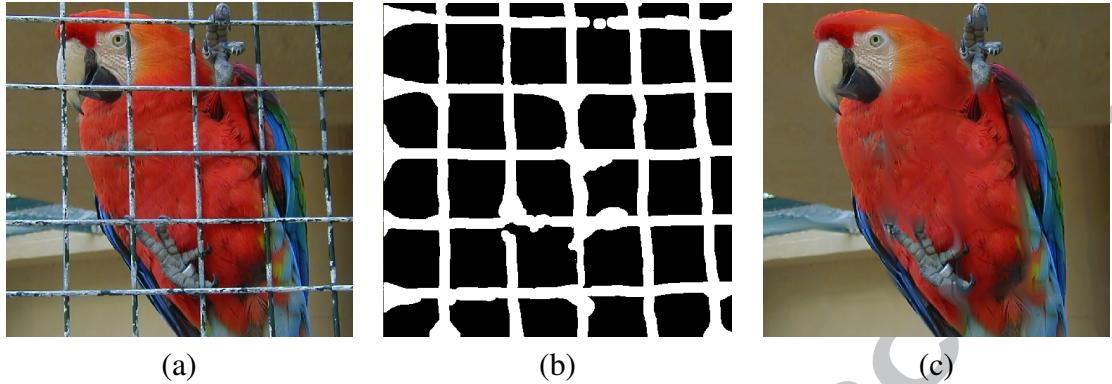


Fig. 1. An example of image inpainting for object removal (a) original image, (b) binary mask, (c) inpainted image [5].



Fig. 2. An example of image inpainting used in restoration, original image (left), restored image (right) [1].

34 In a way, image inpainting can be seen as a modified copy-move tampering process which is
 35 used to recover or remove some parts of the image without any perceptual loss [4, 7]. It is different
 36 from copy-move forgery in the sense that different blocks or regions come from different locations
 37 of the image (see Fig. 5).

38 Although a substantial amount of research has been carried out in developing robust inpaint-
 39 ing algorithms, limited efforts have been put in developing quality assessment metrics to evaluate
 40 the performance of image inpainting (restoration) methods. Image Inpainting Quality Assessment
 41 (IIQA) is a complex and a challenging problem [2]. The objectives of image inpainting assessment
 42 are quite different from those of classical image quality evaluation [9, 10]. Here, for inpainting
 43 IQA, the goal is to evaluate the quality of the restored images. This task can be performed using ei-



Fig. 3. An example of inpainting for photo retouching, original image (left), retouched image (right) [6].

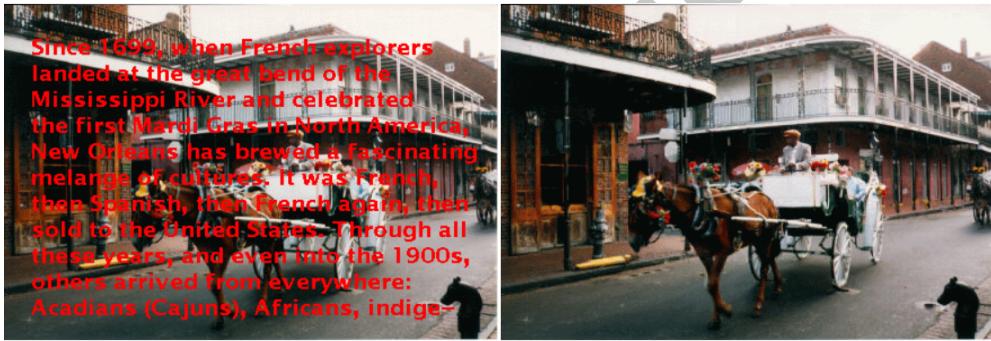


Fig. 4. An example of image inpainting for text removal, original image (left), restored image (right) [1].

ther subjective or objective methods. The main goal here is different, traditional IQA fidelity-based metrics which were mainly developed for quantifying *distortions* in degraded images. Hence, traditional IQA metrics are not well suitable for evaluating the quality of *restored* images and cannot directly be used. This is due to the fact, that restored images in inpainting are different from their original counterparts. Image inpainting is, to some extent, similar to image enhancement. In image enhancement, we start from an input image with poor quality and try to improve its quality. This process is expected to produce more visible structures and the final images are rather different from the original ones. This process also introduces new types of artifacts which affect the perceived quality in a different manner. Among these artifacts, blur is introduced around edges when restoring large inpainted regions. The curved boundaries are not produced correctly as well. Given the unique nature of artifacts introduced in image inpainting and the inappropriateness of

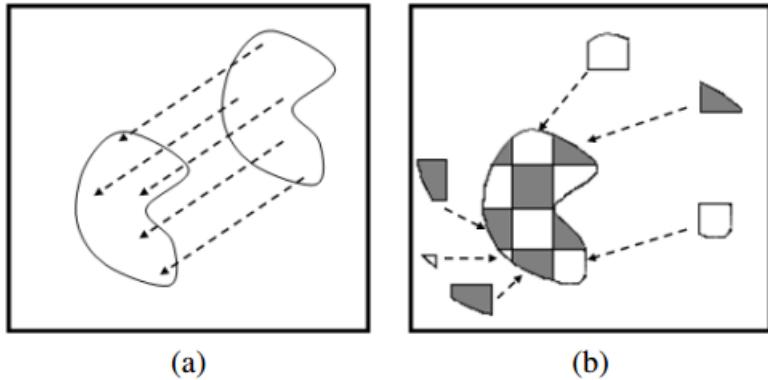


Fig. 5. Difference between two types of tampering (a) copymove forgery, (b) inpainting [8].

55 traditional IQA metrics to quantify quality, we have decided to provide a comprehensive and a
 56 critical review of different methods developed for quality assessment of inpainted images. We
 57 implemented and tested different state-of-the-art IIQA metrics on real images from public image
 58 inpainting database and performance analysis is performed in terms of correlation with the sub-
 59 jective perception-based evaluation. This review will be the first of its kind and is expected to help
 60 researchers working in this area in benchmarking new inpainting techniques, and in developing
 61 more robust methods for inpainting quality assessment, and benchmarking their results.

62 The rest of the paper is structured as follows: Section 2 provides a discussion of common
 63 inpainting algorithms. Section 3 briefly discusses the state-of-the-art IIQA metrics. Different
 64 inpainting databases are discussed in Section 4. The experimental results in terms of correlation
 65 performance are discussed in Section 5. Finally, the paper is concluded in Section 6.

66 2. Image Inpainting Methods

67 The main objective of inpainting algorithms is to restore the unknown regions to create a
 68 more pleasing and realistic feeling about the new (restored) image. Among different types of
 69 inpainting artifacts, most commonly observed ones are blurring, disconnected edges, inconsistent
 70 pieces of texture, etc [1, 2]. Based on our analysis of the state-of-the-art, we propose here to group
 71 inpainting algorithms into four broad categories: Exemplar-based, Partial Differential Equation

72 (PDE)-based, Sparsity-based, and Hybrid (combination of Exemplar-, PDE-, or Sparsity-based)
 73 approaches. We display in Fig. 8 a tree diagram of different classes of most commonly used
 74 inpainting algorithms. Since, the paper aims to provide a critical review of IIQA metrics instead
 75 of the inpainting algorithms themselves, we will only provide a brief discussion of each of these
 76 categories. Note that in image inpainting, the unknown pixels are estimated using the known pixels
 77 information with the assumption that the pixels in the known and unknown regions share similar
 78 geometrical structures and statistical properties.



Fig. 6. An example of disconnected edges in the inpainted image, original image with mask to be filled (left), inpainted image using [2, 11] (right)

79 In PDE-based or diffusion-based inpainting methods, the local structure information is trans-
 80 ferred or diffused from the known region to the unknown (target) region [1]. Several variations of
 81 PDE-based methods were introduced based on the flow of texture information in linear, nonlinear,
 82 isotropic, or anisotropic directions. The PDE-based methods are well-adopted for restoring long
 83 narrow regions (cracks, lines). However, they are not suitable for restoring large unknown tex-
 84 ture regions, due to the introduction of blur in the textured regions (see Fig. 9). The PDE-based
 85 inpainting methods have fast processing time compared.

86 In Exemplar-based inpainting techniques, the structure completion process is carried out using
 87 texture synthesis i.e., the target regions are restored by selecting the patches in the known regions
 88 similar (in terms of structure) to the partially unknown patches in the target regions [4]. *These*

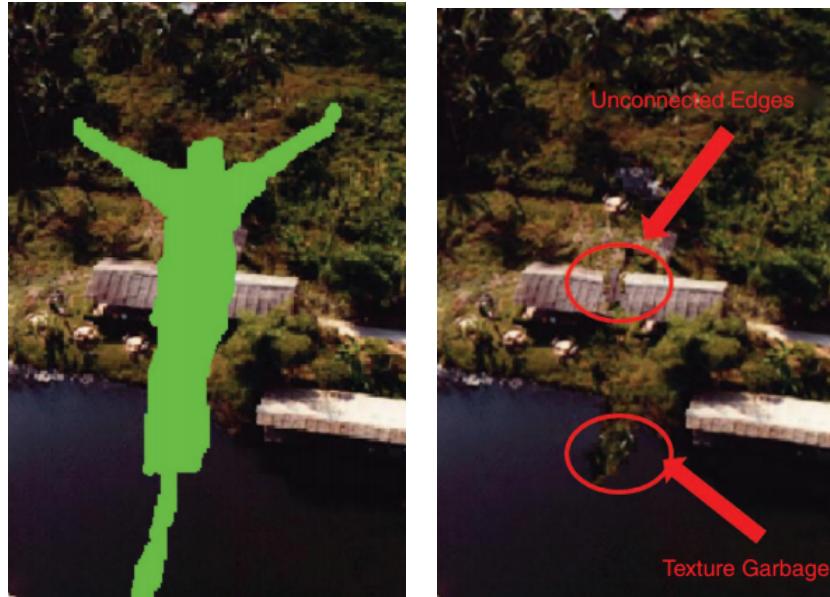


Fig. 7. An example of inconsistent texture regions in exemplar-based methods, original image with mask to be filled (left), inpainted image using [2, 4] (right)

89 techniques use greedy and global optimization cost functions in filling the target regions with
 90 the similar known regions/pixels. In greedy methods, the target region is filled in one iteration
 91 by copying multiple patches in a greedy manner [4, 13]. Whereas, the global optimization and
 92 energy-based techniques minimize a certain energy function in several iterations until convergence
 93 is reached [14].

94 In contrast to PDE-based inpainting methods, Exemplar-based inpainting algorithms perform
 95 better in restoring large regions at the cost of high computation cost. The time required to restore
 96 the unknown patches best matching the known patches, is also high. The choice of patch size is
 97 also critical in Exemplar-based inpainting. For small source patches, it is difficult to find the best
 98 match and results in poor inpainted image quality. Moreover, some undesirable artifacts in terms
 99 of disconnected edges and inconsistent texture regions are visible in the restored images using
 100 exemplar-based inpainting methods. Figs. 6 and 7 show examples of these artifacts produced due
 101 to exemplar-based inpainting.

102 *Sparse representations of images, over a particular basis (e.g., Discrete Cosine Transform*

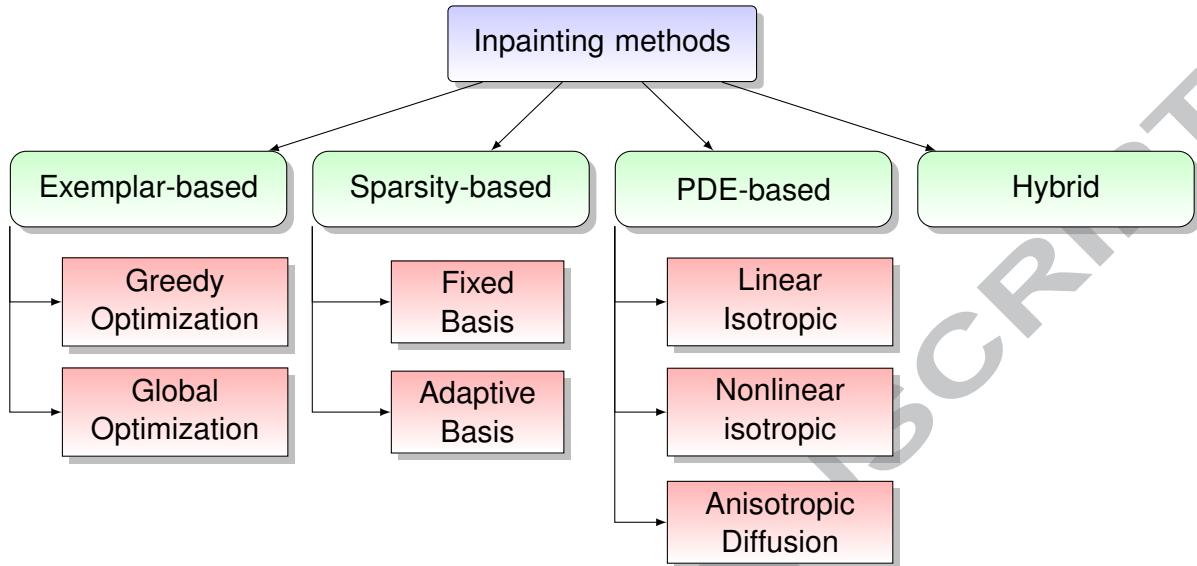


Fig. 8. Tree diagram of different classes of inpainting techniques.

103 (*DCT*), *Discrete Wavelet Transform (DWT)*, etc.) has recently attracted a lot of attraction [15].
 104 Sparsity has also been exploited as a powerful criterion in image inpainting [16, 17]. In Sparsity-
 105 based inpainting methods, the basic idea is representing an image by sparse combination of trans-
 106 formed bases with the assumption that the known and unknown image regions share the same
 107 sparse representations. Therefore, the missing pixels are filled by adaptively updating the sparse
 108 representations. The choice of the dictionary is very important in sparse representations. The
 109 dictionary may be fixed (like *DCT*, *DWT*, etc.) or adaptive/content dependent. In [16], Chan et



Fig. 9. An example of blurring artifacts in PDE-based image inpainting method, original image with large hole to be filled (left), inpainted image using [2, 12] with introduced blur int the large restored regions (right)

¹¹⁰ *al. proposed an inpainting method using the basic Harr-wavelet based fixed dictionary for sparse
¹¹¹ representation of image. Guleryuz et al. in [17] proposed an image inpainting algorithm based on
¹¹² adaptive sparse reconstruction and iterative denoising.*

¹¹³ The Exemplar-based and the Sparsity-based inpainting methods, above, were shown to per-
¹¹⁴ form better the traditional PDE-based methods for filling large texture regions. *Various Hybrid
¹¹⁵ techniques also exist that combine the strengths and different types of inpainting methods for per-
¹¹⁶ formance improvement. These combine PDE-, Exemplar-, and Sparsity-based inpainting methods
¹¹⁷ to reconstruct large missing regions first, then reconstruct missing pixels in the thin regions [18–
¹¹⁸ 20].*

¹¹⁹ To demonstrate the effect of different inpainting algorithms, Fig. 10 shows a very nice ex-
¹²⁰ ample of inpainting where broken pieces of the kiwi fruit are restored using different inpainting
¹²¹ algorithms [21]. The broken area is shown as a green mask. It is clear that Figs. 10 (d) and (f)
¹²² represent more realistic inpainting output compared to the other methods.

¹²³ After this brief survey on commonly used inpainting algorithms, we now move to the focus
¹²⁴ of the paper and discuss in more details different types of subjective and objective IIQA metrics,
¹²⁵ commonly used in the literature.

¹²⁶ **3. Image Inpainting Quality Assessment Measures**

¹²⁷ Image inpainting methods were initially used for removing missing or damaged areas in an
¹²⁸ image [22]. The main criterion was that the restored image should be “close” to the original
¹²⁹ one. The traditional fidelity metrics were used to evaluate the quality of inpainted images. The
¹³⁰ Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR), which are considered as the
¹³¹ most widely used fidelity metrics, were the simplest ones available. Oliveira et al. in [23], for
¹³² example, used these metrics for quality evaluation of inpainted images. However, both MSE and
¹³³ PSNR are not well correlated with perceptual quality assessment [24]. In inpainting applications,
¹³⁴ the objective is to restore the original image such that it is more appealing and that the artifacts

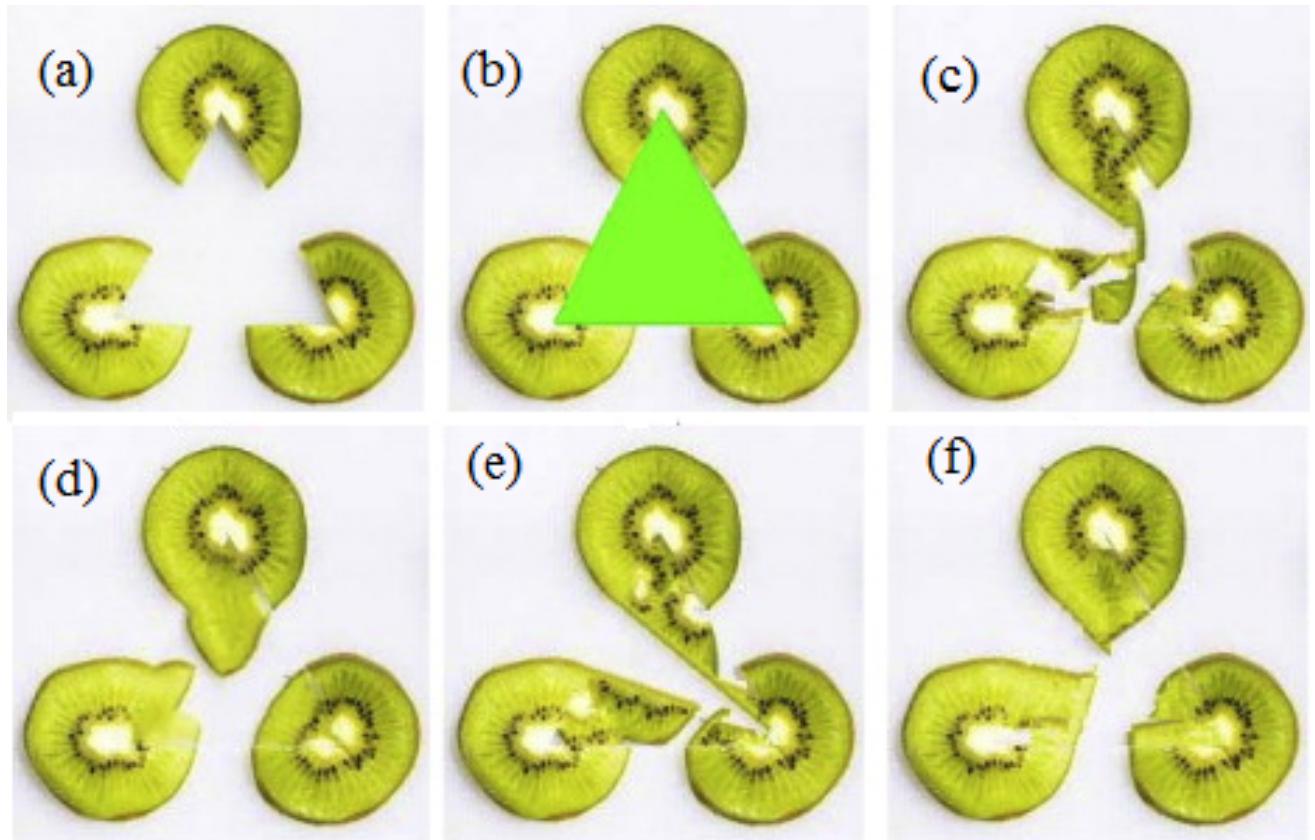


Fig. 10. An example of broken object restoration using inpainting (a) original image, (b) original image with target region in green color, (c-f) inpainting results from different algorithms [21].

135 introduced inside, outside, and around the inpainted regions, are not noticeable/visible.

136 For performance evaluation of different inpainting algorithms, the metric of choice would be a
 137 qualitative judgment averaged over a number of human observers. In this regard, Hays et al. [25]
 138 qualitatively evaluated inpainting image quality for the first time using subjective experiments. The
 139 purpose of the experiment was just the identification of the original and the tampered (inpainted)
 140 images. The proposed method was compared with an exemplar-based approach [4]. Twenty naive
 141 observers participated in the subjective tests and were asked to differentiate between the real image
 142 and the inpainted image (tampered image). The detection rate for tampered images was achieved
 143 as 34%, 64%, and 3% for images in [25], [4], and the original images. The purpose of the study
 144 was to investigate whether or not the proposed inpainted algorithm produced better perceptual

145 quality image compared to other methods. The results showed that the inpainted images from [25]
 146 looked closer to the original images with good perceptual quality. However, the study did not
 147 provide any quantitative ratings of inpainted images.

148 Subjective assessment methods involve humans and the ratings provided are considered as
 149 most reliable and accurate in relation to perceived quality. However, these methods are time-
 150 consuming, laborious, and require a significant number of observers to be consistent. They also
 151 require a well-controlled environment and lighting conditions. This has motivated researchers in
 152 this field to develop alternative objective metrics for inpainting quality assessment without the need
 153 of human involvement. Such objective methods use mathematical tools to extract characteristic
 154 features from either the reference or the test images or both. These features are then used to get a
 155 single quality score for the given image. The aim of objective quality assessment techniques is to
 156 predict perceived image quality, the way a human observer perceives it.

157 Traditional objective quality assessment methods, depending upon the availability of the orig-
 158 inal image, are grouped into Full Reference (FR), Reduced Reference (RR), and No Reference
 159 (NR) methods. In FR methods, the original image is required in addition to the processed image
 160 (inpainted image). These are impractical as the original image is usually not available. With NR
 161 quality prediction methods, the original image is not available. For RR techniques, partial informa-
 162 tion about the original images is available in the form of some extracted features. RR techniques
 163 are seen as a compromise between FR and NR methods. For the case of inpainting, only NR meth-
 164 ods are appropriate due to the unavailability of original image information. Based on our analysis
 165 of the literature, we are proposing to group existing IIQA measures into three broad categories
 166 i.e., structure-based, saliency-based, and machine learning-based measures. We display in Fig. 11
 167 a tree diagram listing the different metrics commonly used for IIQA. The different categories are
 168 now discussed in more details.

169 For the sake of completeness, we outline the notation we used in this paper in Table 1. We
 170 represent an image to be inpainted with I_r and the inpainted image with I_{inp} . The original image to

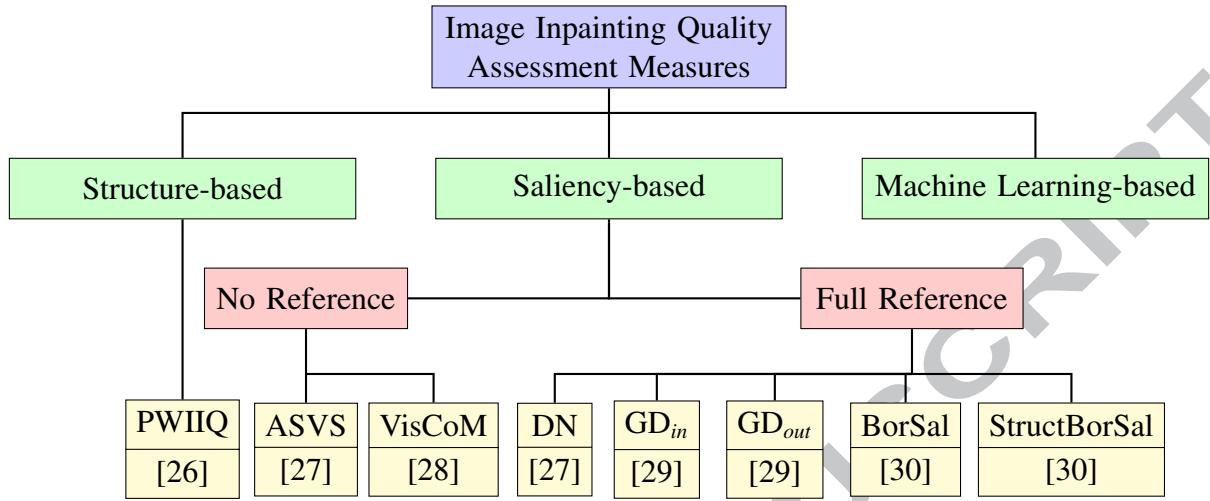


Fig. 11. Proposed framework for grouping the different IIQA metrics.

171 be inpainted is decomposed into three distinct regions, (1) The region to be restored or modified by
 172 the inpainting algorithm is represented by Ω , (2) The remaining area is denoted by Φ , and (3) The
 173 boundary between the two regions is indicated by $\delta\Omega$. Fig. 12a shows a simple image inpainting
 174 model where different regions are clearly labeled.

Table 1
 Notations used for IIQA metrics.

Notation	Description
I_r	Original image
I_{inp}	Inpainted image
$S(.)$	Saliency map of the original image
$S'(.)$	Saliency map of the inpainted image
Ω	Inpainted region
Φ	Remaining region $I - \Omega$
p	Pixel value under consideration
W	Image width
H	Image height
b	Block size
(i, j)	Pixel index

175 After outlining our notation, we will now start discussing each of the groups of the IIQA
 176 metrics mentioned in Fig. 11.

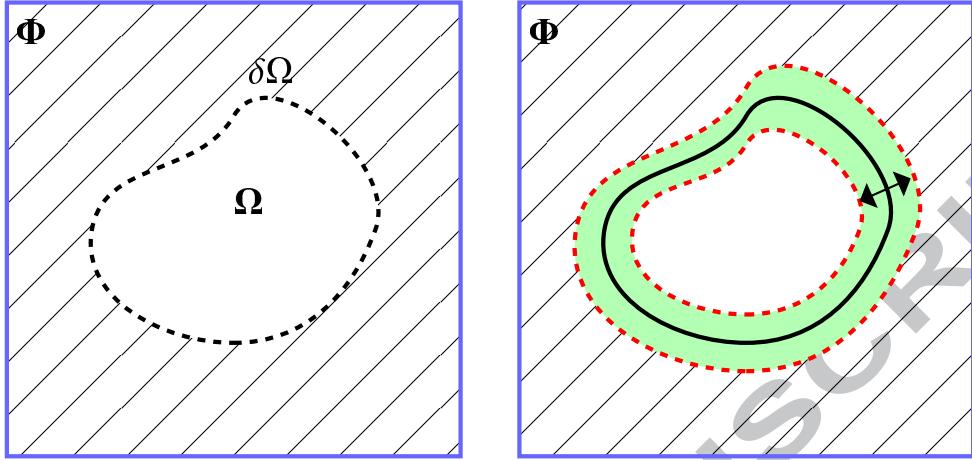


Fig. 12. (a) A simple model used in image inpainting techniques, (b) A typical model used for BorSal and StructBorSal IIQA metrics computations (shaded region is used for IIQA).

177 3.1. Structure-based IIQA measures

178 In image inpainting, some of the structural details in the original image are either removed or
 179 replaced. Inspired by the use of the Structural Similarity Measure (SSIM) [31] in traditional IQA,
 180 Wang et al. [26] proposed a FR metric using luminance, definition, and gradient similarities, to
 181 determine a quality index for inpainted images. The metric, defined as Parameter Weight Image
 182 Inpainting Quality (PWIIQ), is calculated as follows:

$$\text{PWIIQ} = [L(I_r, I_{inp})]^{\alpha} [D(I_r, I_{inp})]^{\beta} [G(I_r, I_{inp})]^{\gamma} \quad (1)$$

183 where the terms L , D , and G , represent the variances of image luminance, definition, and gradient
 184 similarity between the original and inpainted images. The α , β , and γ are positive parameters used
 185 to determine the importance of each term in the final quality score.

186 For implementation purposes, both the original and the inpainted images are first divided into
 187 $b \times b$ fixed-size blocks, and the luminance similarity between the corresponding blocks is com-
 188 puted:

$$l(I_r, I_{inp}) = \frac{2\mu_r\mu_{inp} + K_1}{\mu_r^2 + \mu_{inp}^2 + K_1} \quad (2)$$

189 where μ_r and μ_{inp} represent the mean values of the original and inpainted image blocks respec-

190 tively, while K_1 is a positive constant with very small value to avoid instability when the denominator is close to zero.

192 The weighted block means for the original and inpainted images are used. The weights are
 193 computed from the symmetrical Gaussian filter window of size 11×11 pixels. The resulting
 194 weighted mean is given by:

$$\mu = \sum_{i=1}^N w_i x_i \quad (3)$$

195 where $w = \{w_i \text{ such that } \sum_{i=1}^N w_i = 1, \quad i = 1, 2, \dots, N\}$, where N denotes the number of pixels
 196 in the window.

197 The luminance component, L , in equation (1) is computed as the average of the luminance
 198 similarities across all blocks:

$$L(I_r, I_{inp}) = \frac{1}{B_1 \times B_2} \sum_{i=1}^{B_1} \sum_{j=1}^{B_2} l(I_r^{ij}, I_{inp}^{ij}) \quad (4)$$

199 where B_1 and B_2 represent the number of overlapping blocks along the rows and columns of the
 200 image.

201 Secondly, the image definition function, D , is computed as follows:

$$D(I_r, I_{inp}) = \frac{\sum_{i=0}^{W-1} \sum_{j=0}^{H-1} |\mathcal{F}_{inp}^{ij}| - |\mathcal{F}_{inp}^{00}|}{\sum_{i=0}^{W-1} \sum_{j=0}^{H-1} |\mathcal{F}_r^{ij}| - |\mathcal{F}_r^{00}|} \quad (5)$$

202 where $\mathcal{F}(.)$ represents the Fourier transform of an image and \mathcal{F}_{00} is the dc component or overall
 203 mean value of an image.

204 Finally, the gradient similarity component is defined as:

$$G(I_r, I_{inp}) = \frac{2 \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} G_r^{ij} G_{inp}^{ij} + K_2}{\sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [G_r^{ij}]^2 + \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [G_{inp}^{ij}]^2 + K_2} \quad (6)$$

205 where the $G(.)$ represents the gradient magnitude computed from the Sobel filter mask of size 3×3

206 in the vertical and horizontal directions and K_2 is a small positive constant.

207 Similarly to its IQA counterpart, the structure-based methods suffer from some serious limi-
 208 tations. Since image inpainting operations do not require the original images, the large inpainted
 209 regions may be quite different from the actual ones. Consequently, the structural similarity based
 210 methods (e.g., [26]) may fail for images with large inpainted regions (see Fig. 9). To overcome the
 211 drawbacks of structure-based methods, researchers started introducing image saliency to derive
 212 new measures for evaluating quality of inpainted images. Other variants of the SSIM have also
 213 been proposed with limited success [32].

214 *3.2. Saliency-based IIQA measures:*

215 Visual saliency plays a significant role in image quality assessment applications. Image saliency
 216 is used to highlight the areas towards which the human vision system is more sensitive/attracted.
 217 Various saliency detection algorithms exist in the literature [33–35]. Given its importance in IQA,
 218 saliency has been used in estimating visibility of different artifacts introduced by inpainting. The
 219 basic idea is that salient regions change after inpainting. The most prominent IIQA metrics using
 220 the concept of saliency are now briefly outlined.

221 *3.2.1. Average Squared Visual Salience (ASVS)*

222 In [27], Ardis et al. proposed two objective metrics for quality assessment of inpainted images.
 223 The image saliency was used in capturing the distortions introduced during the restoration process.
 224 The first metric is the ASVS, which is represented by the normalized sum of squares of the saliency
 225 values within the inpainted region. The ASVS metric relates to the noticeability of the inpainted
 226 pixels compared to the overall scene. ASVS is a NR metric as it does not require the original
 227 image information. It is calculated as follows:

$$\text{ASVS} = \frac{1}{\|\Omega\|} \sum_{p \in \Omega} |S'(p)|^2 \quad (7)$$

228 where $S'(p)$ represents the saliency value for the inpainting pixel, p , within to the inpainted region,
 229 Ω . High values of the ASVS correspond to more visibility of inpainting related artifacts and

230 reduced perceptual quality [27].

231 *3.2.2. Degree of Noticeability (DN)*

232 Ardis et al. in [36], categorized inpainting artifacts into two broad classes, i.e., in-region and
 233 out-region artifacts. During the restoration operation in image inpainting, the in-region artifacts
 234 occur due to the introduction of distinct color and structures in the inpainted regions only. These
 235 artifacts result in an increased saliency in the inpainted areas and thus disturb attention flow. The
 236 ASVS metric relates to the in-region artifacts as it only considers the salient pixels within the
 237 inpainted region.

238 Similarly, the out-region artifacts occur when the local colors or structures are not properly
 239 extended to the inpainted region by the inpainted method. These artifacts result in an increase in
 240 the saliency of the inpainted region neighborhood and hence decreases attention flow within the
 241 inpainted region. The in-region and out-region artifacts are computed as follows:

$$\text{In-region} = \text{ASVS} = \frac{1}{\|\Omega\|} \sum_{p \in \Omega} |S'(p)|^2 \quad (8)$$

$$\text{Out-region} = \frac{1}{\|\Phi\|} \sum_{p \in \Phi} |S'(p) - S(p)|^2 \quad (9)$$

242 243 Ardis et al., in [27], took into account both in-region and out-region artifacts and proposed
 244 another metric, namely the DN (Degree of Noticeability). The DN measure is intended to identify
 245 non-noticeable inpainted regions and indicates the change in attention flow in the neighborhood of
 246 the inpainted regions. It is calculated as follows:

$$\text{DN} = \frac{\|\Omega\|}{\|\Omega\| + \|\Phi\|} \text{in-region} + \frac{\|\Phi\|}{\|\Omega\| + \|\Phi\|} \text{out-region} \quad (10)$$

247 Equation (10) can further be simplified as follows:

$$\text{DN} = \frac{1}{\|\Omega\| + \|\Phi\|} \left(\sum_{p \in \Omega} |S'(p)|^2 + \sum_{p \in \Phi} |S'(p) - S(p)|^2 \right) \quad (11)$$

248 For both ASVS and DN calculations, the saliency maps are generated using the iLab Neu-

romorphic Vision Toolkit (iNVT) version 3.1, using scale-4 and discretization of 1 : 16. The expected visual cortex stimulation was set with 0.1 ms observation cutoff. Furthermore, four orientation scales, three center scales (2 to 4), and two center-surround channels (3,4) were considered [27].

Similarly to the ASVS, high values of the DN correspond to more visibility of inpainting related artifacts and reduced perceptual quality. The authors claimed a good correlation for both metrics, with subjective ratings. However, the subjective ratings were not considered reliable as only three observers participated in the psychophysical experiment. Moreover, the overall visual appearance of an image is also ignored while calculating DN and ASVS IIQA metrics.

3.2.3. Gaze Density (GD)-based IIQA measures

Following the work in [27], Mahalingam et al. [29] proposed two visual saliency-based metrics for inpainting quality assessment within and outside the inpainted regions. From eye-tracking experimental data, the gaze densities were used to capture the saliency in the original and inpainted images. The motivation was that changes in the saliency map in the inpainted image is related to its perceptual quality.

In their subjective experiments, 45 reference images and 90 modified images were obtained using two different inpainting algorithms. The images were equally distributed into three subsets. Twenty-four naive observers without any prior knowledge of the original and the inpainted images rated the subgroups under ambient lighting conditions and at a distance of 65cm from the display screen. The average gaze distribution was calculated for each image from the eye-tracking experiment. It was observed that the Human Visual System (HVS) is more attracted towards the regions with more noticeable inpainting artifacts (see Fig. 13). The gaze density was calculated for both inside and outside the inpainted regions using:

$$GD_{in} = \frac{1}{\|\Omega\|} \sum_{p \in \Omega} S'(p) \quad (12)$$

272

$$\text{GD}_{\text{out}} = \frac{1}{\|\Phi\|} \sum_{p \in \Phi} S'(p) \quad (13)$$

273 The gaze density measures of the inpainted image were normalized by the gaze densities of the
 274 original image to account for variations in textures and sizes. The final normalized gaze densities
 275 were given by:

$$\overline{\text{GD}}_{\text{in}} = \frac{\sum_{p \in \Omega} S'(p)}{\sum_{p \in \Omega} S(p)}, \quad \text{and} \quad (14)$$

$$\overline{\text{GD}}_{\text{out}} = \frac{\sum_{p \in \Phi} S'(p)}{\sum_{p \in \Phi} S(p)} \quad (15)$$

The experimental results showed a strong correlation between the rankings from the subjective

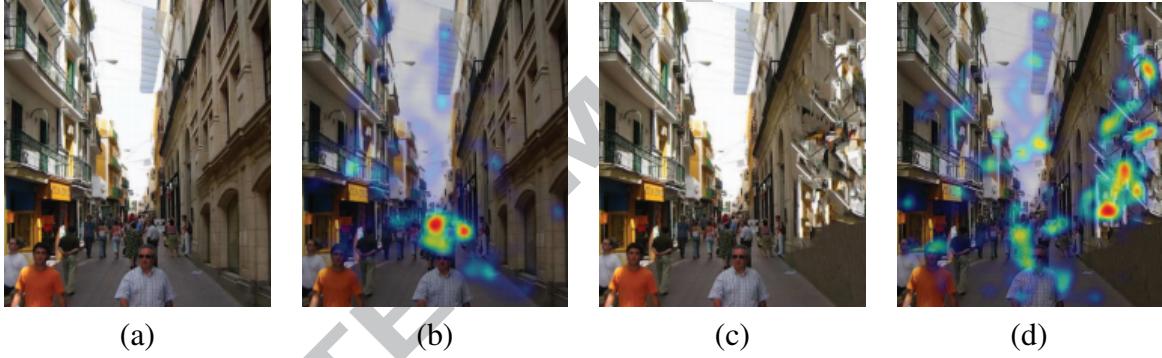


Fig. 13. An example of concentration of gaze distribution or HVS attention around discernible artifacts regions [29]
 (a) Image inpainted using [25], (b) Average gaze distribution on (a), (c) Image inpainted using [4] having visible artifacts, and (d) Average gaze distribution on (c).

277
 278 experiments and the gaze density based measures. However, these methods require the original
 279 image and are not suited for practical inpainting applications, where the original image is usually
 280 unavailable. Similar to the ASVS and the DN metrics, both GD_{in} and GD_{out} highlight the change in
 281 attention flow within and outside the inpainted regions respectively, and do not consider the global
 282 visual appearance of the image. However, these measures are computationally more efficient com-
 283 pared to the ASVS and DN, since summation is carried over the salient pixels in the expressions
 284 of the GD-based IIQAs, rather than the square summations in the ASVS and DN IIQA measures.

285 3.2.4. *Border Saliency based Measures (BorSal)*

286 The measures proposed in [27, 29] considered the in-region and out-region artifacts separately.
 287 Oncu et al. [30] showed that saliency map pixels in the neighborhood of the inpainted region are
 288 sufficient to capture the changes in saliency due to inpainting. The BorSal metric was proposed
 289 to compute the normalized gaze density using the border pixels extended only to three pixels
 290 inside and outside the inpainted regions. The six pixels wide border area simultaneously contain
 291 information from both in-region and out-region artifacts (see Fig. 12b). The BorSal metric was
 292 computed as follows:

$$\text{BorSal} = \frac{\sum_{p \in \text{Border}} S'(p)}{\sum_{p \in \text{Border}} S(p)} \quad (16)$$

293 3.2.5. *Structural Border Saliency based Measures (StructBorSal)*

294 The BorSal IIQA metric accounts for changes in the flow of attention over the inpainted im-
 295 age. Oncu et al. [30] proposed another metric called StructBorSal, to account for the structure
 296 information in the whole image and to highlight the artifacts in the restored image. The SSIM_{IPT}
 297 based measure [37] was used by taking the geometrical mean of the three SSIM computed for each
 298 color channel separately. The StructBorSal metric combines the BorSal metric with the structural
 299 measure as follows:

$$\text{StructBorSal} = \text{BorSal} + \text{SSIM}_{IPT} \quad (17)$$

300 The correlations between the subjective ratings and 14 quality measures (IQA metrics for distor-
 301 tions as well as inpainting) were calculated. The results showed poor performance of existing
 302 metrics. The inpainting IQA metrics performed well for images with small and less structured
 303 inpainted regions.

304 The above mentioned saliency-based inpainting IQA metrics, i.e., DN [27], GD_{in} [29], GD_{out}
 305 [29], BorSal [30], and StructBorSal [30] require the original image whereas in the restoration
 306 process, usually the original image is not available. The overall visual appearance of the image
 307 also plays a significant role in the quality perception. These metrics are also lacking in considering

308 the global visual appearance of an image, limiting their use in practical setups.

309 *3.2.6. Visual Coherence Metric (VisCoM)*

310 In image inpainting, the reference image is usually not available, therefore the restored pix-
 311 els rely solely on the surrounding pixels. The restored pixels in the inpainted regions, however,
 312 should exhibit consistency with existing pixels. The coherence of the inpainted regions, which
 313 is associated with the degree of annoyance of noticeable distortions, is computed by taking the
 314 correlation between the inpainted pixels and the existing ones in the region outside the inpainted
 315 region. Similarly, the HVS is more sensitive to the edges and contours in an image. The presence
 316 of contours and edge details are more attracted by the HVS compared to the remaining regions.
 317 The saliency map of a given image relates to the degree of attention in the image and hence can be
 318 used to weight the coherence map in evaluating the final quality index.

319 Trung et al. [28, 38–40], proposed a NR quality metric using visual coherence and visual
 320 saliency of restored regions. The index is computed as follows:

$$\text{VisCoM} = \frac{1}{\|\Omega\|} \sum_{p \in \Omega} C(p)^\alpha S(p)^\beta \quad (18)$$

321 where $C(p)$ and $S(p)$ represent the coherence term and the saliency or structure term respectively.

322 The exponents α and β control the significance of each term in the final quality score.

323 The coherence term, C , is basically a similarity index between the inpainted regions and the
 324 remaining ones in the inpainted image. It is defined as follows:

$$C(p) = \max [SIM(\Psi_p, \Psi_q), \quad \text{all } \Psi_q \in \Phi \quad \text{and} \quad \text{all } \Psi_p \in \Omega] \quad (19)$$

325 where Ψ_p and Ψ_q represent small patches around pixels, p and q, respectively. The SIM is the
 326 similarity function between two patches. Instead of the MSE or PSNR, the SSIM index [31], used
 327 in classical IQA, is exploited here to find the coherence between patches of size 7×7 . The SIM is
 328 defined as:

$$SIM(\Psi_p, \Psi_q) = \frac{(2\mu_p\mu_q + K_1)(2\sigma_{pq} + K_2)}{(\mu_p^2 + \mu_q^2 + K_1)(\sigma_p^2 + \sigma_q^2 + K_2)} \quad (20)$$

329 where μ_p , σ_p and μ_q , σ_q represent the mean and standard deviation of the patches, Ψ_p and Ψ_q ,
 330 centered at pixels, p and q, respectively, whereas σ_{pq} is the cross correlation between the patches
 331 Ψ_p and Ψ_q , and K_1 and K_2 are small positive coefficients to insure stability when denominator is
 332 zero.

333 The local structure term is computed from the saliency values which are further used as weights
 334 in the final quality index. Among different saliency detection algorithms, the authors used a sim-
 335 ple and computationally efficient method for salient region detection [41]. In [41], color and
 336 luminance information were used for saliency detection. For a given image, I , the saliency map
 337 was generated using:

$$S' = \|I_\mu - I_G\| \quad (21)$$

338 where I_μ represents the mean value of the original image and I_G is a Gaussian blurred (5×5 filter
 339 mask) version of the original image. The operation is performed in the CIELab color space. The
 340 method is simple, computationally efficient, and does not need any downsampling operation during
 341 the estimation of the saliency map. Finally, the saliency map, defined in Eq. 21, is normalized to
 342 the range [0, 1]:

$$S(p) = \frac{S'(p)}{\max_I(S')} \quad \forall p \in \Omega \quad (22)$$

343 The authors in [28, 38–40] used visual coherence of recovered regions and visual saliency
 344 describing visual importance to develop their index shown in Eq. 18. The proposed approach
 345 showed promising results but could only handle a limited number of possible inpainting artifacts.

346 Based on our study of existing approaches, we present in Table 2 a summary of both structure-
 347 based and saliency-based IIQA metrics. It is important to note that there exists only two NR-
 348 IIQA metric among these metrics. To overcome this limitation, among others, researchers tried
 349 to use advanced machine learning approaches in developing robust quality assessment metrics for
 350 practical inpainting applications.

Table 2
A summary of structure-based and saliency-based IIQA measures.

Metric	Year	Type	Regions used	Description	Strengths	Weaknesses
PWIIQ[26]	2008	FR	overall image	statistical features used. A value close to 1 means better quality	simple and fast	requires reference image, fails for images with large holes (inpainted regions).
ASVS[27]	2009	NR	in-region	used when fidelity is not important. High value means more visibility of artifacts, poor quality	reference image is not needed	ignore overall visual appearance of an image
DN [27]	2009	FR	in-region, out-region	used for applications where preservation of original saliency is desired.	highlights change in attention flow beyond the inpainted regions	requires reference image, ignore overall visual appearance of an image
GD _{in} [29]	2010	FR	in-region	value close to 1 means no deviation of attention flow in the inpainted image	highlights change in attention flow within the inpainted regions	requires reference image,
GD _{out} [29]	2010	FR	out-region	same as GD _{in}	highlights change in attention flow beyond the inpainted regions	requires reference image,
BorSal [30]	2012	FR	border region	a single border region around the hole (inpainted region) is used.	fast, a single border region around the hole (inpainted region) is used.	requires reference image,
StructBorSal [30]	2012	FR	border region	in addition to BorSal, structural information is also used.	fast, a single border region around the hole (inpainted region) is used, structural artifacts are also captured	requires reference image, coherence of the inpainted regions with remaining is ignored
VisCoM [28]	2013	NR	overall image	uses visual coherence along with structural information	overall global visual appearance of image is considered	requires reference image,

351 *3.3. Machine Learning-based IIQA Measures*

352 Machine learning-based approaches were originally developed for solving classification and
 353 regression problems efficiently and provide good approximation of functional relationships be-
 354 tween input features and output classes/scores scores from the training session. In the testing
 355 stage, a set of features is extracted from a given image. The trained model and the extracted
 356 features are then used for predicting the quality rating of the test image [42].

357 Among the first approaches using machine learning for IIQA is the metric proposed by Viach-
 358 eslav et al. in [43]. The method is a NR approach for IIQA based on natural scene statistics and
 359 machine learning. First, the saliency map of the inpainted image is calculated to identify most
 360 important perceptual information in the inpainted image. The saliency map is then thresholded
 361 using average gaze density computed from the outside inpainted regions using Eq. (13) for proto-
 362 objects. Then, the DCT is calculated only for the proto-objects and used to train a dictionary of
 363 100 classes, where each word in the dictionary is a DCT coefficient. For each DCT block, the his-
 364 togram of words is then used as a feature vector. The quality scores collected from the subjective
 365 experiments and the extracted features were then used to train a Support Vector Regression (SVR)
 366 network, and to predict the quality of inpainted images resulting from different algorithms.

367 The same authors in [44], replaced the DCT based features with the traditional Local Binary
 368 Pattern (LBP) features given their power in describing image structures effectively. The quality
 369 scores collected from the subjective experiments and the extracted features are then used to train
 370 an SVR for quality prediction. For the subjective experiments, a database consisting of 300 images
 371 with different structures and textures was used. The database also included some real images. The
 372 images were restored using a mask and using four different inpainted methods. Ten observers
 373 participated in the subjective experiments and rated the quality of the inpainted images on a scale
 374 1-5 (5 for excellent quality, 1 for worst quality). The results showed good correlation with human
 375 ratings of quality.

376 More recently, Markio et al. [45] showed that saliency is not an absolute requirement for

377 assessing inpainting quality. They performed an experiment using a learning-to-rank approach.
 378 Instead of determining the absolute scores for inpainted images, the preference order is obtained
 379 among inpainted images from different inpainting algorithms. They demonstrated that visual
 380 saliency map is useful but not a requirement. Rather, they showed that some features can be used
 381 to reflect the changes within and outside the modified areas in an inpainted image. Such features
 382 are extracted from gaze measurements using a simple Tobii eye-tracker device. From each original
 383 image, twelve inpainted images are generated using two inpainted methods, three patch sizes, and
 384 two multiscale parameters. A hundred and eleven original images were used in the experiments.
 385 The proposed metric was compared to other existing metrics in terms of prediction accuracy in
 386 estimating the preferences order ranking. The authors showed that existing saliency-based IIQA
 387 metrics fail in ordering the inpainted images correctly due to the small significant difference in
 388 the saliency maps in the inpainted regions. The results using the proposed metric showed an im-
 389 provement of at least 7% over other metrics with 68.65% prediction accuracy. To summarize the
 390 work on IIQA using machine learning techniques, we present in Table 3, a brief summary of most
 391 prominent approaches.

392 Before leaving the above discussion on IIQA, we now discuss another important issue of high
 393 relevance to IQA applications; that is the collection of quality ratings averaged over a number of
 394 observers over comprehensive databases with inpainted images from different algorithms. These
 395 databases are used as a ground truth for validating and testing or both inpainting algorithms and
 396 IIQA metrics.

397 **4. Image Inpainting Quality Assessment Databases**

398 With the tremendous increase of research activities in image inpainting algorithms and ap-
 399 plications, it was crucial to develop comprehensive databases for performance evaluation and
 400 benchmarking of different inpainting methods. In the literature, usually, the performance of
 401 an inpainting algorithm is evaluated on own local images or using standard databases used for

Table 3
A summary of machine learning-based IIQA measures.

Method	Year	Type	Feature Description	Regression
Voronin et al.[43, 46]	2014	NR	DCT-based dictionary	SVR, RBF ¹ kernel
Voronin et al.[44]	2015	NR	LBP histograms	SVR, EMD ² kernel
Markio et al.[45]	2016	NR	Gaze features	RankingSVM, RBF kernel

¹ RBF (Radial Basis Function)

² EMD (Earth Mover's Distance)

402 standard IQA problems. Given the importance of image inpainting in multimedia applications,
 403 publicly-available databases are needed for unbiased performance comparison. In this regards,
 404 Tiefenbacher et al. provided, for the first time, a public database namely the Technische Uni-
 405 versitt Mnchen Image Inpainting Database (TUM-IIID) [47], for objectively estimating quality of
 406 inpainted images and performance evaluation of different IIQA metrics. The database contained
 407 17 reference images with diverse texture types and resolution of 640×480 pixels stored in PNG
 408 format. Each image in the database is inpainted using four state-of-the-art inpainted methods [48–
 409 51] and for four inpainting regions. Then, each inpainted image in the database was rated by
 410 21 observers using a Single Stimulus (SS) subjective experiment protocol, and the ratings from
 411 all observers were averaged to get a single score for each image. Some sample images, inpaint-
 412 ing masks, and inpainted images from public and private databases are shown in Figs. 14 and 15
 413 respectively.

414 In an effort to summarize existing work in inpainting using different databases (private and
 415 public), we also present in Table 4, the most common experimental setups used in the literature.

416 5. Experimental Results and Discussions

417 The main objective of the paper is to provide a comprehensive performance evaluation and
 418 critical review of different state-of-the-art IIQA metrics. Therefore, after providing a detailed de-
 419 scription of the existing IIQA metrics, the IIQA metrics are investigated for their consistency with

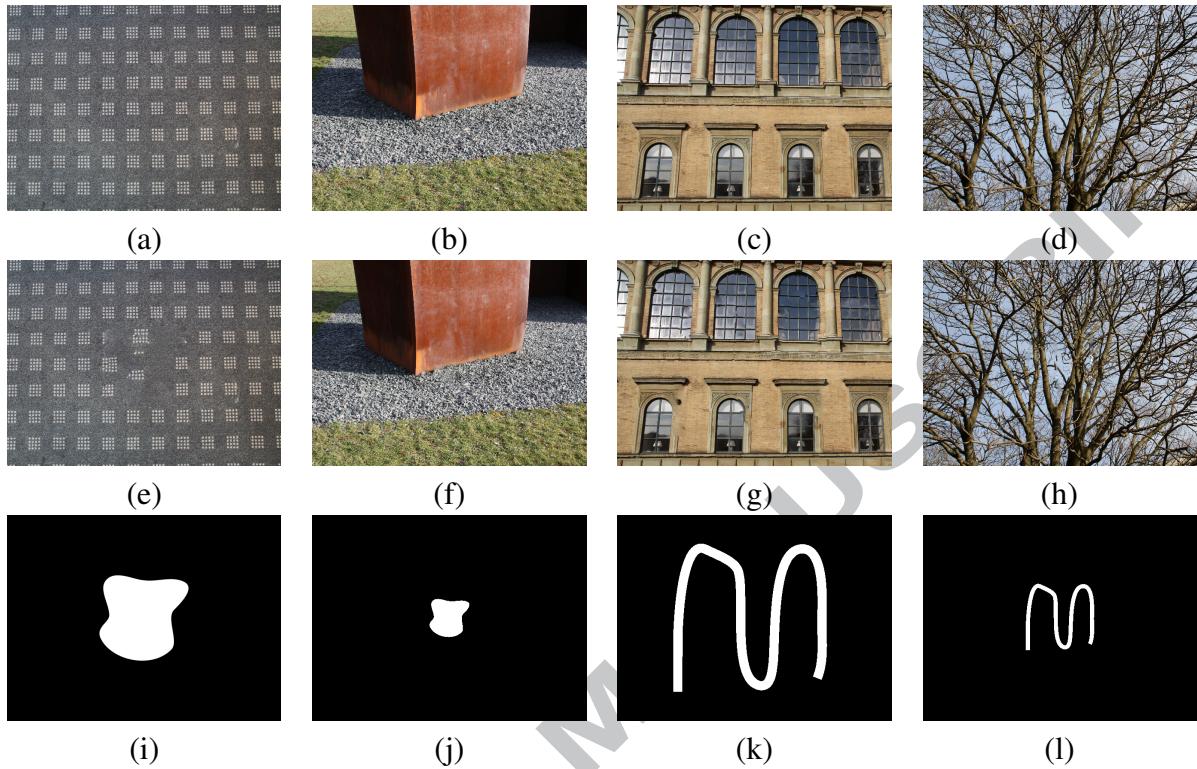


Fig. 14. Sample images from the TUM-IID [47] database: (a-d) Reference images, (e-h) inpainted images using [48], (i-l) masks used for inpainting [48].



Fig. 15. Sample local images used in [40] for inpainting quality assessment (a) original image with mask, inpainted image using [4] (b), [54] (c), and [55] (d).

Table 4
Experimental setups used for private and public inpainting databases.

Method	[26]	[29]	[30]	[27]	[56]	[47]	[40]	[45]
Year	2008	2009	2010	2012	2013	2015	2015	2016
Original images	3	-	45	6	8	300	7	111
Inpainted images	9	-	90	48	-	1200	56	2466
Image resolution	-	-	-	-	-	-	640 × 480	-
Subjective method ¹	rank-order	-	SS	DS	web	SS	SS	rank
Score type	ranking	4-scale	gaze density	5-scale	5-scale	5-scale	5-scale	5-scale
Screen Resolution ¹	-	20cm ht.	1280 × 960	-	-	-	-	1280 × 1080
Observers	-	5	24	69	15 – 20	10	21	24
Viewing distance ³	-	0.7cm	65cm	web	-	-	-	60cm
Inpainting methods	3	-	2	8	-	4	4	2
Evaluation measures	2	2	2	14	-	4	3	6
Access	private	private	private	private	private	private	public	private

¹ SS (Single Stimulus), DS (Double Stimulus)

² ht. (height)

³ web (through web-based interface)

420 the human visual perception [52]. For this purpose, we have used the psychophysical subjective
 421 experimental data provided with the TUMIID [47], a public image inpainting database. For per-
 422 formance evaluation, we have used the Spearman Rank Order Correlation Coefficient (SROCC)
 423 measure. The SROCC is widely used as a non-parametric measure to determine the monotonicity
 424 between the ranks of two variables and its value ranges from -1 to $+1$. The value is close to $+1$
 425 in the case of strong correlation between the ranks of two variables, and -1 in the case of strong
 426 disagreement between the two variables. The SROCC gives zero value when there is no correla-
 427 tion between the ranks. In our study, the aim is to observe how well an IIQA measure is consistent
 428 in capturing the ranking for the four inpainted versions of each original image in the database.
 429 Therefore, before evaluating the correlations, we must consider, how the change in the magnitude
 430 of metric values affects image quality. For some metrics, high values correspond to good quality,
 431 whereas for other metrics, the opposite is true (see Section 3). For preference ranking, the highest
 432 score is highly ranked. Whereas, the metrics with high/low values corresponding to good quality
 433 are also highly ranked. Using the concept of strength of correlation alone without signs does not
 434 convey the desired purpose [53].

435 It has been shown that the performance of an IQA measure is substantially related to certain
 436 image characteristics [57]. Therefore, in this work, we calculate the image-wise correlations for
 437 the selected IIQA metrics. For each image in the database, we have the ranking scores for its four
 438 inpainted versions as well as the quantitative scores.

439 If I_i represents an original image and $I_{i,j}$, its inpainted version processed by method M_j , for
 440 $i = 1, 2, \dots, n_I$ and $j = 1, 2, \dots, n_J$, for $(n_I = 7, n_J = 4)$. Here n_I and n_J represent the number of
 441 original images and the number of inpainting methods respectively. We compute the SROCC for
 442 each image using the following relation:

$$\rho_{i,k} = 1 - \frac{6d_{i,j}^2}{n_J(n_J^2 - 1)}, \text{ for } i = 1, 2, \dots, n_I \quad (23)$$

443 where $d_{i,j}$ represents the difference in the ranks of subjective preferences and objective scores of

Table 5

Image-wise correlation performance of the IIQA metrics using the compact inpainting masks.

image	2	3	6	7	10	14	17	Mean SROCC
ASVS [27]	-1	-1	-1	-0.8	-1	-0.8	-0.8	-0.9143
DN [27]	-1	-1	-1	-0.8	-1	-0.8	-0.8	-0.9143
GDin [29]	1	1	0.8	0.8	1	0.8	0.8	0.8857
GDout [29]	0.4	0.2	0.4	-0.4	-0.4	0.8	0	0.1429
BorSal [30]	0.8	1	0.8	0.8	0.4	0.8	0.4	0.7143
StructBorSal [30]	0.8	1	1	0.8	0.4	0.8	0.4	0.7429
VisCOM [28]	0.8	1	0.8	0.4	0.4	1	0.4	0.6857
PWIIQ [26]	0.8	1	0.8	0	-0.8	0.8	0.8	0.4857

444 the k^{th} IIQA measure for the i^{th} image. The image-wise correlation and mean SROCC's for two
 445 different types of inpainting masks (i.e., compact and spread-out masks) are reported in Tables 5
 446 and 6.

447 From the results, it is clear that the performance of the metrics is effected by the choice of the
 448 inpainted regions. For compact inpainted regions, it is difficult to restore the missing pixels com-
 449 pared to the spread-out (thin) inpainted regions. Obviously, evaluating quality when reconstructing
 450 large regions is more challenging than when only small regions are missing due to long term mem-
 451 ory requirements for large regions. From the results, it is clear that IIQA metrics showed betetr
 452 correlation performance in the case of spread-out (thin) inpainted regions compared to the com-
 453 pact (large) inpainted regions. The ASVS and the DN metrics showed strong disagreements with
 454 the subjective ratings (negative correlation values) for both types of inpainted regions. Whereas,
 455 other metrics provide good correlation performance with the subjective judgements. The GDin,
 456 BorSal, StructBorSal, and VisCOM IIQA metrics showed consistent correlation performance for
 457 both types of inpainting regions.

458 *5.1. Computational Time Analysis*

459 In real-time applications, low-complexity IIQA metrics are needed. Therefore, in addition to
 460 comparing the correlation performance, we also compute the computational time of each IIQA
 461 metric. We list in Table 7, the time taken (in seconds) to compute the metric score for a single im-

Table 6

Image-wise correlation performance of the IIQA metrics using the spread-out inpainting masks.

image	2	3	6	7	10	14	17	Mean SROCC
ASVS [27]	1	-1	-0.8	-0.4	-1	-1	-1	-0.8857
DN [27]	-1	-0.8	-0.8	-0.4	-1	-1	-1	-0.8571
GDin [29]	1	1	0.8	0.4	1	1	0.4	0.8000
GDout [29]	1	1	0.8	0.4	1	1	0.4	0.8000
BorSal [30]	0.8	1	0.8	0.8	0.4	1	0.8	0.8000
StructBorSal [30]	0.8	1	0.6	1	0.4	1	0.4	0.7429
VisCOM [28]	1	1	0.4	0.8	0.8	0.4	0.6	0.7143
PWIIQ [26]	0.8	1	1	0.8	0.4	1	1	0.8571

Table 7

Execution time for different IIQA metrics.

Metric	Time (seconds)
ASVS [27]	0.3897
DN [27]	0.8471
GDin[29]	0.8943
GDout [29]	0.8657
BorSal[30]	0.8541
StructBorSal[30]	2.845
VisCOM[28]	300
PWIIQ [26]	0.2436

- age of resolution 512 x 512. The experiments were performed on a notebook Intel Core i5-24500M CPU@2.5GHz and 4G RAM. The software platform is MATLAB R2013b under Windows 8.1. The execution times are only provided to show the relative complexity of the different metrics. The results show that among the NR IIQA metrics, the VisCOM which is good in terms of correlation requires the largest computational time while the ASVS requires second least computational time but has poor correlation performance. Whereas, among various FR-IIQA metrics, the BorSal, and StructBorSal have moderate computational complexity and good correlation performance. The PWIIQ metric provides the least (best) computational time with moderate correlation performance. The results in Table 7 only provide an idea about the comparative complexity of different IIQA metrics. In real-time applications, the algorithms can be substantially optimized.

472 **6. Conclusion and Future trends**

473 The quality assessment of inpainted images continues to be a complex and challenging prob-
 474 lem. It is substantially different, in nature, from classical IQA, due to the different types of artifacts
 475 not commonly observed in other applications, the importance of perceptual clues in scene conti-
 476 nuity, and the need for convincing and plausible recovered images. While the advances made in
 477 developing robust methods for inpainting are significant, little efforts have been put in developing
 478 inpainting-dedicated quality assessment (IIQA) metrics. To bridge this gap, we present this survey
 479 paper which summarizes current research efforts carried in the development of robust IIQA mea-
 480 sures as well as current challenges and future research directions and applications of inpainting
 481 and inpainting-related quality assessment measures. We also provide a performance comparison
 482 of different metrics in terms of correlation performance and computational complexity.

483 Our study revealed that among existing IIQA metrics, many of these require the availability
 484 of the original image, whereas image inpainting is usually used in the case of unavailability of
 485 the reference or original image. The need for developing robust NR-IIQA metrics is still per-
 486 sistent. From the review of state-of-the-art, we also noticed that saliency-based methods fail for
 487 cases where no difference in perceptual saliency can be seen around the damaged/restored regions.
 488 Machine learning-based are shown to overcome some of these limitations. Moreover, rather than
 489 determining the absolute quality rating scores, the preference order estimation-based methods
 490 were shown to provide very convincing results.

491 We also observed that most of the developed metrics were designed and validated over private
 492 databases consisting of a limited number of images and human raters. Currently, only one public
 493 database [47] exists which contains a limited number of images. Given the importance of this
 494 evolving field in multimedia, it is becoming crucial to develop new public databases consisting of
 495 large numbers of inpainted images, generated using various inpainted methods. This will help in
 496 providing unbiased comparisons among different IIQA metrics, highlighting their shortcomings,
 497 and in introducing new efficient quality measures well-correlated with human perception of quality

498 of inpainting operations.

499 The overall performance of IIQA measures also depends on the selected features, and how
 500 these are exploited in scene continuity and assessing resulting artifacts. The features describing
 501 naturalness, colorfulness, continuity of pixels, among others, around the vicinity of damaged re-
 502 gions were shown to facilitate the IIQA task.

503 Our analysis also showed the importance of understanding image content before selecting the
 504 most appropriate IIQA metric. We have seen that while the majority of inpainting techniques
 505 focus on small regions continuity (e.g., PDE), others work better with large regions (e.g., Exem-
 506 plar). Obviously, evaluating quality when reconstructing large regions is more challenging than
 507 when only small regions are missing due to long term memory requirements for large regions. As
 508 such, IIQA metrics should be designed (or selected) taking into consideration whether the given
 509 inpainting image contains large or small reconstructed regions.

510 Currently, we see a growing number of inpainting applications embedded in new generation
 511 mobile devices to remove (and reconstruct) certain objects from captured photos. Hence, com-
 512 putational efficiency of inpainting and IIQA algorithms becomes more relevant to these types of
 513 platforms. Other applications of IIQA are the quality prediction of inpainted images over the cloud
 514 or in wireless environments. Video inpainting is also a challenging problem when there is a need
 515 to remove and track undesired objects in videos or movies. Research in Video Inpainting Quality
 516 Assessment (VIQA) is expected to flourish over the next few years as we see more people working
 517 in this all-important branch of multimedia.

518 References

- 519 [1] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester, Image inpainting, in: Proceedings of the 27th annual confer-
 520 ence on Computer graphics and interactive techniques - SIGGRAPH '00, ACM Press, New York, USA, 2000,
 521 pp. 417–424. doi:10.1145/344779.344972.
- 522 [2] C. Guillemot, O. Le Meur, Image Inpainting: Overview and Recent Advances, IEEE Signal Processing Maga-
 523 zine 31 (1) (2014) 127–144. doi:10.1109/MSP.2013.2273004.
- 524 [3] Z. Chen, C. Dai, L. Jiang, B. Sheng, J. Zhang, W. Lin, Y. Yuan, Structure-aware image inpainting using
 525 patch scale optimization, Journal of Visual Communication and Image Representation 40 (2016) 312–323.
 526 doi:10.1016/j.jvcir.2016.06.029.

- 527 [4] A. Criminisi, P. Perez, K. Toyama, Region Filling and Object Removal by Exemplar-Based Image Inpainting,
 528 IEEE Transactions on Image Processing 13 (9) (2004) 1200–1212. doi:10.1109/TIP.2004.833105.
- 529 [5] M. A. Qureshi, M. Deriche, A bibliography of pixel-based blind image forgery detection techniques, Signal
 530 Processing: Image Communication 39 (2015) 46–74. doi:10.1016/j.image.2015.08.008.
- 531 [6] Inpaint photo restoration software.
 532 URL <http://www.theinpaint.com/>
- 533 [7] Z. Liang, G. Yang, X. Ding, L. Li, An efficient forgery detection algorithm for object removal by exemplar-
 534 based image inpainting, Journal of Visual Communication and Image Representation 30 (2015) 75–85.
 535 doi:10.1016/j.jvcir.2015.03.004.
- 536 [8] D. T. Trung, A. Beghdadi, M.-C. Larabi, Blind inpainting forgery detection, in: 2014 IEEE Global
 537 Conference on Signal and Information Processing (GlobalSIP), IEEE, Atlanta, GA, USA, 2014, pp. 1019–1023.
 538 doi:10.1109/GlobalSIP.2014.7032275.
- 539 [9] A. Beghdadi, M. C. Larabi, A. Bouzerdoum, K. M. Iftekharuddin, A survey of perceptual image processing
 540 methods, Signal Processing: Image Communication 28 (8) (2013) 811–831. doi:10.1016/j.image.2013.06.003.
- 541 [10] M. Qureshi, A. Beghdadi, B. Sdiri, M. Deriche, F. A. Cheikh, A Comprehensive Performance Evaluation of Ob-
 542 jective Quality Metrics For Contrast Enhancement Techniques, in: European Workshop on Visual Information
 543 Processing (EUVIP), 2016, pp. 1–5.
- 544 [11] O. Le Meur, C. Guillemot, Super-Resolution-Based Inpainting, in: European Conference on Computer Vision
 545 (ECCV), Florence, Italy, 2012, pp. 554–567. doi:10.1007/978-3-642-33783-3_40.
- 546 [12] A. Telea, An Image Inpainting Technique Based on the Fast Marching Method, Journal of Graphics Tools 9 (1)
 547 (2004) 23–34. doi:10.1080/10867651.2004.10487596.
- 548 [13] O. Le Meur, J. Gautier, C. Guillemot, Examplar-based inpainting based on local geometry, in: Image Processing
 549 (ICIP), 2011 18th IEEE International Conference on, IEEE, 2011, pp. 3401–3404.
- 550 [14] P. Arias, V. Caselles, G. Facciolo, Analysis of a variational framework for exemplar-based image inpainting,
 551 Multiscale Modeling & Simulation 10 (2) (2012) 473–514.
- 552 [15] M. A. Qureshi, M. Deriche, A new wavelet based efficient image compression algorithm using compressive
 553 sensing, Multimedia Tools and Applications 75 (12) (2016) 6737–6754.
- 554 [16] R. H. Chan, S. Setzer, G. Steidl, Inpainting by flexible haar-wavelet shrinkage, SIAM Journal on Imaging
 555 Sciences 1 (3) (2008) 273–293.
- 556 [17] O. G. Guleryuz, Nonlinear approximation based image recovery using adaptive sparse reconstructions and iter-
 557 ated denoising-part i: theory, IEEE Transactions on image processing 15 (3) (2006) 539–554.
- 558 [18] N. Komodakis, G. Tziritas, Image completion using efficient belief propagation via priority scheduling and
 559 dynamic pruning, IEEE Transactions on Image Processing 16 (11) (2007) 2649–2661.
- 560 [19] F. Cao, Y. Gousseau, S. Masnou, P. Pérez, Geometrically guided exemplar-based inpainting, SIAM Journal on
 561 Imaging Sciences 4 (4) (2011) 1143–1179.
- 562 [20] M. Bertalmio, L. Vese, G. Sapiro, S. Osher, Simultaneous structure and texture image inpainting, IEEE Trans-
 563 actions on image processing 12 (8) (2003) 882–889.
- 564 [21] R. Rodriguez-Sánchez, J. A. García, J. Fdez-Valdivia, Image inpainting with nonsubsampled contourlet trans-
 565 form, Pattern Recognition Letters 34 (13) (2013) 1508–1518. doi:10.1016/j.patrec.2013.06.002.
- 566 [22] I. Peterson, Filling in Blanks, Science News 161 (19) (2002) 299–300. doi:10.2307/4013521.
- 567 [23] M. M. Oliveira, B. Bowen, R. McKenna, Y.-S. Chang, Fast Digital Image Inpainting, in: Proceedings of the
 568 International Conference on Visualization, Imaging and Image Processing (VIIP), Marbella, Spain, 2001, pp.
 569 261–266.
- 570 [24] Zhou Wang, Bovik, Ligang Lu, Why is image quality assessment so difficult?, in: IEEE Interna-
 571 tional Conference on Acoustics Speech and Signal Processing, Vol. IV, IEEE, 2002, pp. 3313–3316.
 572 doi:10.1109/ICASSP.2002.1004620.
- 573 [25] J. Hays, A. A. Efros, Scene completion using millions of photographs, Communications of the ACM 51 (10)
 574 (2008) 87. doi:10.1145/1400181.1400202.
- 575 [26] S. Wang, H. Li, X. Zhu, P. Li, An Evaluation Index Based on Parameter Weight for Image Inpainting Quality,
 576 in: 2008 The 9th International Conference for Young Computer Scientists, IEEE, Hunan, 2008, pp. 786–790.
 577 doi:10.1109/ICYCS.2008.461.

- 578 [27] P. A. Ardis, A. Singhal, Visual salience metrics for image inpainting, in: M. Rabbani, R. L. Stevenson
 579 (Eds.), Proc. SPIE 7257, Visual Communications and Image Processing, San Jose, CA, 2009, p. 72571W.
 580 doi:10.1117/12.808942.
- 581 [28] T. T. Dang, A. Beghdadi, M.-C. Larabi, Visual coherence metric for evaluation of color image restoration,
 582 in: 2013 Colour and Visual Computing Symposium (CVCS), IEEE, Gjovik, Norway, 2013, pp. 1–6.
 583 doi:10.1109/CVCS.2013.6626268.
- 584 [29] V. V. Mahalingam, S.-c. S. Cheung, Eye tracking based perceptual image inpainting quality analysis,
 585 in: 2010 IEEE International Conference on Image Processing, IEEE, Hong Kong, 2010, pp. 1109–1112.
 586 doi:10.1109/ICIP.2010.5653640.
- 587 [30] A. I. Oncu, F. Deger, J. Y. Hardeberg, Evaluation of Digital Inpainting Quality in the Context of Artwork
 588 Restoration, in: Proceeding of the 12th international conference on computer vision, 2012, pp. 561–570.
 589 doi:10.1007/978-3-642-33863-2_58.
- 590 [31] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to struc-
 591 tural similarity, *IEEE Transactions on Image Processing* 13 (4) (2004) 600–612. doi:10.1109/TIP.2003.819861.
- 592 [32] T.-J. Liu, Y.-C. Lin, W. Lin, C.-C. J. Kuo, Visual quality assessment: recent developments, coding appli-
 593 cations and future trends, *APSIPA Transactions on Signal and Information Processing* 2 (e4) (2013) 1–20.
 594 doi:10.1017/AT SIP.2013.5.
- 595 [33] X. Ma, X. Xie, K.-m. Lam, J. Hu, Y. Zhong, Saliency detection based on singular value decomposition, *Journal*
 596 of *Visual Communication and Image Representation* 32 (2015) 95–106. doi:10.1016/j.jvcir.2015.08.003.
- 597 [34] X. Ma, X. Xie, K.-M. Lam, Y. Zhong, Efficient saliency analysis based on wavelet transform
 598 and entropy theory, *Journal of Visual Communication and Image Representation* 30 (2015) 201–207.
 599 doi:10.1016/j.jvcir.2015.04.008.
- 600 [35] Q. Zhao, C. Koch, Learning saliency-based visual attention: A review, *Signal Processing* 93 (6) (2013) 1401–
 601 1407.
- 602 [36] P. A. Ardis, C. M. Brown, A. Singhal, Inpainting quality assessment, *Journal of Electronic Imaging* 19 (1) (2010)
 603 011002. doi:10.1117/1.3267088.
- 604 [37] N. Bonnier, F. Schmitt, H. Brettel, S. Berche, Evaluation of Spatial Gamut Mapping Algorithms, in: 14th Color
 605 and Imaging Conferenc (CIC), 2006, pp. 56–61. doi:10.1.1.497.9528.
- 606 [38] D. T. Trung, A. Beghdadi, C. Larabi, Perceptual quality assessment for color image inpainting, in: 2013
 607 IEEE International Conference on Image Processing (ICIP), IEEE, Melbourne, VIC, 2013, pp. 398–402.
 608 doi:10.1109/ICIP.2013.6738082.
- 609 [39] D. T. Trung, A. Beghdadi, M.-c. Larabi, Perceptual Evaluation of Digital Image Completion Quality, in: 21st
 610 European Signal Processing Conference (EUSIPCO 2013), Marrakech, 2013, pp. 1–5.
- 611 [40] T. T. Dang, A. Beghdadi, M. C. Larabi, U. Paris, U. D. Poitiers, Inpainted image quality assessment, in: Visual
 612 Information Processing (EUVIP), 2013 4th European Workshop on, Paris, France, 2013, pp. 76–81.
- 613 [41] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: 2009
 614 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Miami, USA, 2009, pp. 1597–1604.
 615 doi:10.1109/CVPR.2009.5206596.
- 616 [42] M. A. Qureshi, M. Deriche, A fast no reference image quality assessment using laws texture moments, in: 2014
 617 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, Atlanta, GA, USA, 2014,
 618 pp. 979–983. doi:10.1109/GlobalSIP.2014.7032267.
- 619 [43] V. Viacheslav, F. Vladimir, M. Vladimir, G. Nikolay, S. Roman, F. Valentin, Low-level features for inpaint-
 620 ing quality assessment, in: 2014 12th International Conference on Signal Processing (ICSP), Vol. 53, IEEE,
 621 Hangzhou, 2014, pp. 643–647. doi:10.1109/ICOSP.2014.7015082.
- 622 [44] V. Voronin, V. Marchuk, E. Semenishchev, S. Maslennikov, I. Svirin, Inpainted image quality assessment based
 623 on machine learning, in: WSCG 2015 Conference on Computer Graphics, Visualization and Computer Vision,
 624 2015, pp. 167–172.
- 625 [45] M. Isogawa, D. Mikami, K. Takahashi, A. Kojima, Eye gaze analysis and learning-to-rank to obtain the most
 626 preferred result in image inpainting, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE,
 627 Phoenix, Arizona, USA, 2016, pp. 3538–3542. doi:10.1109/ICIP.2016.7533018.
- 628 [46] V. A. Frantc, V. V. Voronin, V. I. Marchuk, A. I. Sherstobitov, S. Agaian, K. Egiazarian, Machine learn-

- 629 ing approach for objective inpainting quality assessment, in: S. S. Agaian, S. A. Jassim, E. Y. Du (Eds.),
630 Proc. SPIE 9120, Mobile Multimedia/Image Processing, Security, and Applications, 2014, p. 91200S.
631 doi:10.1117/12.2063664.
- 632 [47] P. Tiefenbacher, V. Bogischef, D. Merget, G. Rigoll, Subjective and objective evaluation of image inpainting
633 quality, in: IEEE Intl.Conf. on Image Processing (ICIP), IEEE, Quebec City, Canada, 2015, pp. 447–451.
- 634 [48] A. Bugeau, M. Bertalmio, V. Caselles, G. Sapiro, A Comprehensive Framework for Image Inpainting, IEEE
635 Transactions on Image Processing 19 (10) (2010) 2634–2645. doi:10.1109/TIP.2010.2049240.
- 636 [49] Z. Xu, J. Sun, Image inpainting by patch propagation using patch sparsity, IEEE Transactions on Image Pro-
637 cessing 19 (5) (2010) 1153–1165.
- 638 [50] P. Getreuer, Total variation inpainting using split bregman, Image Processing On Line 2 (2012) 147–157.
- 639 [51] J. Herling, W. Broll, Pixmix: A real-time approach to high-quality diminished reality, in: International Sympo-
640 sium on Mixed and Augmented Reality, IEEE, 2012, pp. 141–150.
- 641 [52] S. Siegel, Nonparametric Statistics for The Behavioral Sciences, 2nd Edition, McGraw-Hill, 1956.
- 642 [53] M. Rubinstein, D. Gutierrez, O. Sorkine, A. Shamir, A comparative study of image retargeting, ACM Transac-
643 tions on Graphics 29 (6) (2010) 1. doi:10.1145/1882261.1866186.
- 644 [54] Q. Zhang, J. Lin, Exemplar-based image inpainting using color distribution analysis, Journal of information
645 science and engineering 28 (4) (2012) 641–654.
- 646 [55] T.-T. Dang, C. Larabi, A. Beghdadi, Multi-resolution patch and window-based priority for digital image inpaint-
647 ing problem, in: 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA),
648 IEEE, 2012, pp. 280–284.
- 649 [56] V. V. Voronin, V. A. Frantc, V. I. Marchuk, A. I. Sherstobitov, K. Egiazarian, Inpainted image quality assessment
650 based on machine learning (report), in: K. O. Egiazarian, S. S. Agaian, A. P. Gotchev (Eds.), Proceedings of the
651 SPIE, 2015, p. 93990U. doi:10.1117/12.2076507.
- 652 [57] J. Y. Hardeberg, E. Bando, M. Pedersen, Evaluating colour image difference metrics for gamut-mapped images,
653 Coloration Technology 124 (4) (2008) 243–253.

Manuscript Number: JVCI-17-5R1

Title: A critical survey of state-of-the-art image inpainting quality assessment metrics

Highlights:

We present a thorough overview of image inpainting quality assessment (IIQA) metrics.

We introduce a new framework for clustering IIQA metrics into major groups.

We provide a comprehensive performance analysis of IIQA metrics on public databases.

We outline the strengths and weaknesses of existing IIQA metrics.

We discuss future research directions on IIQA metrics and their applications.