



HAL
open science

Computational Analysis of Correlations between Image Aesthetic and Image Naturalness in the Relation with Image Quality

Quyét-Tien Le, Patricia Ladret, Huu-Tuan Nguyen, Alice Caplier

► **To cite this version:**

Quyét-Tien Le, Patricia Ladret, Huu-Tuan Nguyen, Alice Caplier. Computational Analysis of Correlations between Image Aesthetic and Image Naturalness in the Relation with Image Quality. *Journal of Imaging*, 2022, 8 (6), pp.166. 10.3390/jimaging8060166 . hal-03976328

HAL Id: hal-03976328

<https://hal.science/hal-03976328v1>

Submitted on 7 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

Computational Analysis of Correlations between Image Aesthetic and Image Naturalness in the Relation with Image Quality

Quy-tien LE ¹ , Patricia LADRET ², Huu-Tuan NGUYEN ¹ and Alice CAPLIER ²

¹ Faculty of Information Technology, Vietnam Maritime University, 484 Lach Tray, Le Chan, 04000 Hai Phong, VIETNAM; tienlqcnt@vamaru.edu.vn; huu-tuan.nguyen@vamaru.edu.vn

² GIPSA Lab, Grenoble Alpes University, 11 rue des Mathématiques, Grenoble Campus BP 46, F-38402 Saint Martin d'Heres Cedex FRANCE; patricia.ladret@gipsa-lab.grenoble-inp.fr; alice.caplier@gipsa-lab.grenoble-inp.fr

* Correspondence: tienlqcnt@vamaru.edu.vn

Version March 9, 2022 submitted to J. Imaging

Abstract: The main purpose of this paper is the study of the correlations between Image Aesthetic (IA) and Image Naturalness (IN) and the analysis of the influence of IA and IN on Image Quality (IQ) in different contexts. The first contribution is a study about the potential relationships between IA and IN. For that study, two sub questions are considered. The first one is to validate the idea that IA and IN are not correlated to each other. The second one is about the influence of IA and IN features on Image Naturalness Assessment (INA) and Image Aesthetic Assessment (IAA) respectively. Secondly, it is obvious that IQ is related to IA and IN but the exact influence of IA and IN on IQ has not been evaluated. Beside that, the context impact on those influences has not been clarified so the second contribution is to investigate the influence of IA and IN on IQ in different contexts. Results obtained from rigorous experiments prove that although there are moderate and weak correlations between IA and IN, they are still two different components of IQ. It also appears that viewers' IQ perception is affected by some contextual factors and the influence of IA and IN on IQ depends on the considered context.

Keywords: image quality; image aesthetic; image naturalness; human visual perception; image quality assessment; image aesthetic assessment; image naturalness assessment; quality of experience

1. Introduction

In recent years, there are more and more personal devices integrated with digital cameras such as smartphones, tablets, laptops. It leads to a dramatic increase of the number of photos day by day so users' storage tends to be filled very fast. Therefore, evaluating photos to keep the best ones and to remove the worst ones becomes an essential need. This task is traditionally performed based on human visual system. Figure 1 shows an overview of image factors affecting human visual perception. Those factors are categorized into two groups: image content and Image Quality (IQ). On one side, image content obviously has a great influence on human visual perception. This group contains three factors: "message inside", "emotional inspiration" and "image originality". In Figure 1(1), the first photo is an example for "message inside" with a bird stuck in a plastic bag. Although the content looks simple, it contains a hidden message related to environment like "let's save animals", "stop consuming plastic bags" or "our planet is destroyed". Regarding Figure 1(2), some people might not have any special feelings about the photo but the hug between the bride and her father could remind other people of their family members or a personal memory. The value of the photo is "emotional inspiration". In Figure 1(3), a rare moment of a cloudy sky with light rays makes the photo different from other

31 photos of the same landscape. Although there is no hidden message or emotional inspiration in this case, the originality makes the photo special. On the other side, image perception might be based

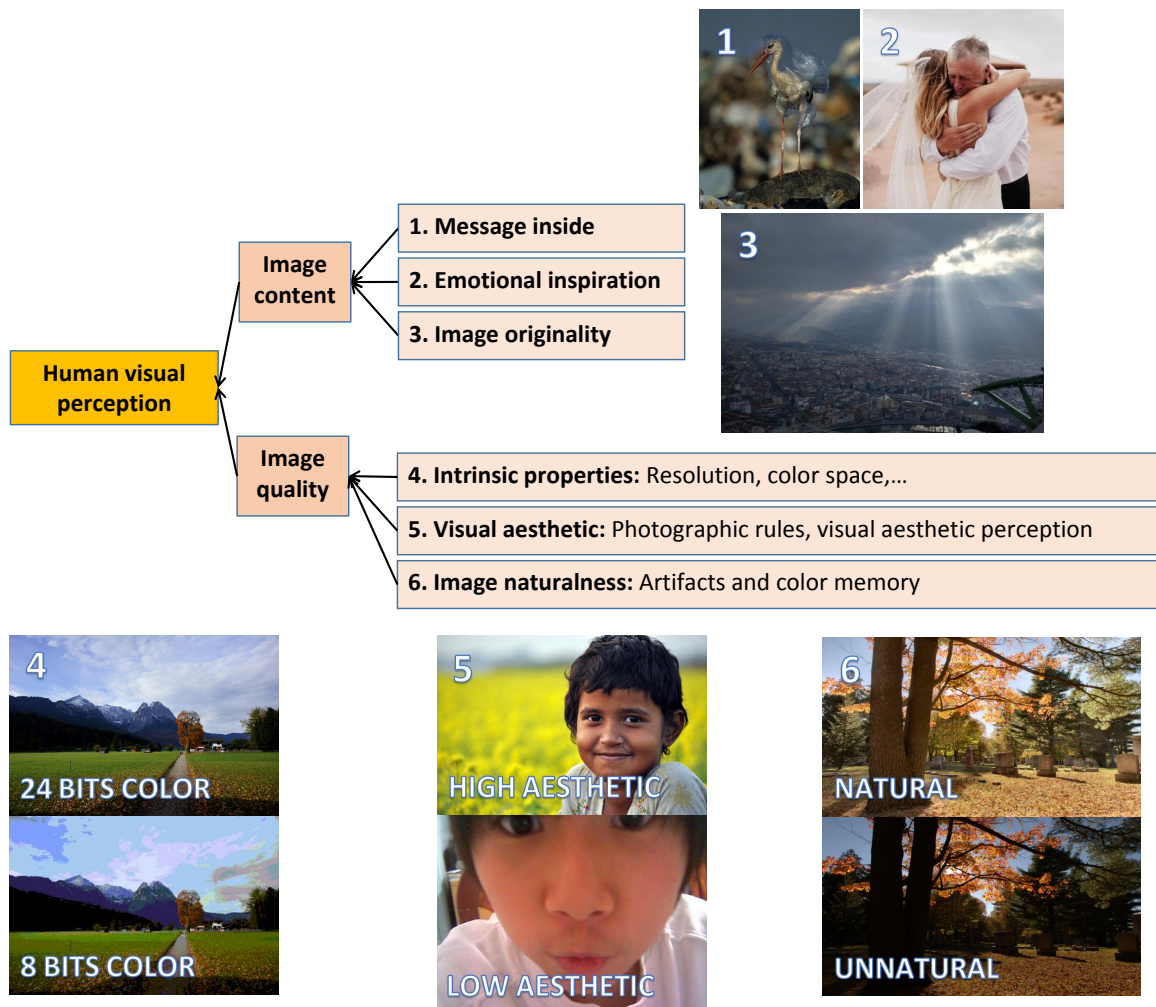


Figure 1. Overview of image aspects having influence on human visual perception.

32
 33 on Image Quality (IQ) which is not supposed to be related to image content. In this paper, we are
 34 focusing on IQ only. IQ in this study is defined in an explicit way. It is generally based on three
 35 notions: intrinsic properties, Image Aesthetic (IA - this work considers this notion related to visual
 36 aesthetic only, other aspects related to image content are not considered in this study) and Image
 37 Naturalness (IN). "Intrinsic properties" is a notion related to some technical aspects of photos such as
 38 resolution, color space, color depth, image format (see Figure 1(4) and Figure 2). This notion mostly
 39 refers to properties of the image acquisition device and it is not related to any external factors induced
 40 by viewers' experience or contexts. In the past, intrinsic properties were the main factor influencing
 41 IQ since acquisition device performances were not so good (low resolution, optical or chromatic
 42 aberrations). However, this has been reduced because of the improvement of the technology so the role
 43 of intrinsic properties in IQ is less and less significant. That is the reason why in our study, we make
 44 the assumption that intrinsic properties do not influence IQ on the image databases we are dealing
 45 with.

46 On the contrary, IA is the measure of how aesthetically a photo fulfills the observer's expectation
 47 based on photography rules and individual visual aesthetic perception (see Figure 1(5) and Figure 3(a)).
 48 This notion is related to what happens in viewers' mind when they look at a photo.



Figure 2. Examples of intrinsic properties. The first photo has a higher resolution than the others while the color depth of the third photo is shallower than those of the two first ones.

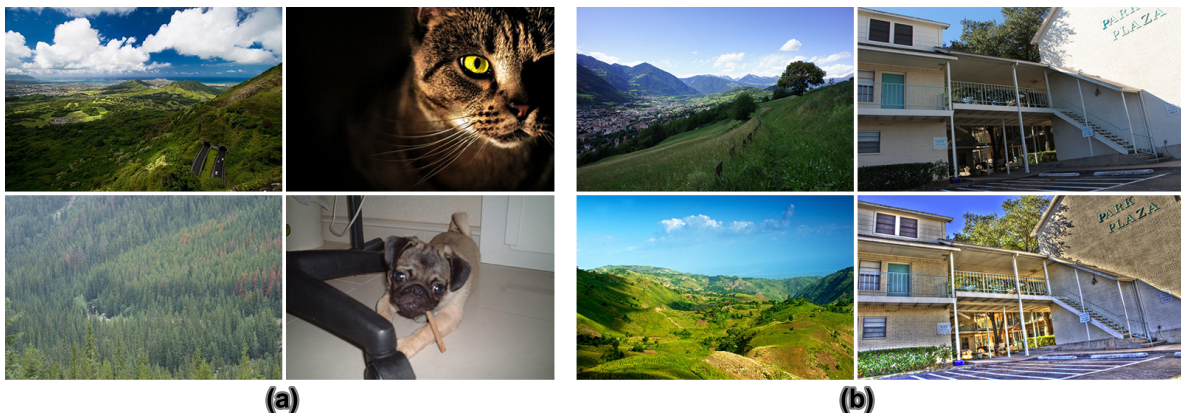


Figure 3. (a) IA illustrations: examples of high aesthetic images are in the first row while the second row contains examples of low aesthetic images. (b) IN illustrations: examples of natural images are in the first row while the second row contains examples of unnatural images.

49 On the side of IN, this notion is both related to artifacts induced by some image processing
50 algorithms and to the individual feeling [1] (see Figure 1(6) and Figure 3(b)). Regarding artifacts,
51 IN is affected by strong visible clues detected by viewers' eyes so the unnaturalness feeling comes
52 from annoying artifacts induced by camera sensors, image processing algorithms (compressing,
53 tone-mapping), image format, file transfer (see bottom right photo in Figure 3(b)). In contrast, the
54 feeling of naturalness and unnaturalness might also come from viewer's experience and memory (see
55 bottom left photo in Figure 3(b)). When viewing a photo, observers compare the scene in the photo to
56 the reality retrieved from their memory (what they have seen) to find differences and similarities, so
57 the feelings are not the same for all viewers. There is also a part of subjectivity in IN perception.

58 To sum up, this paper focuses on subjective aspects of IQ including IA and IN. Intrinsic properties
59 are not supposed to influence IQ in our study. Although there are many researches about IQ, IA and
60 IN, the potential relationships between IA, IN and IQ are still an open question. The main purpose
61 of this study is to clarify the correlations between IA and IN - two subjective aspects of IQ and to
62 investigate the relations between contextual factors and the impact of IA and IN on IQ. Understanding
63 those correlations might be a potential base to simulate human visual perception and enhance IQ based
64 on IA and IN. Although many studies of IA and IN have been introduced for decades, a ground-truth
65 with both IA and IN data from subjective experiments does not exist. Therefore, instead of using
66 subjective data, this work is approached computationally. IA and IN perception is modeled based on
67 subjective ground-truth of IA and IN separately. By using those simulated models, IA and IN might
68 be measured. The simulated data is then used to analyze the correlations between IA and IN in the
69 relation with IQ.

70 2. State of the Art

71 2.1. Image Quality

72 IQ can be approached from different angles. One common way is to implicitly define IQ with
73 regard to a given Image Quality Assessment (IQA) protocol: either Full Reference Image Quality
74 Assessment (FRIQA), No Reference Image Quality Assessment (NRIQA) or reduced reference IQA.
75 On the side of FRIQA, there are many purposes for image transformation such as image compression,
76 tone mapping, steganography, image enhancement. It is assumed that the chain of operations induces
77 negative effects on the quality of the transformed versions compared to the quality of the original
78 version. Therefore, IQ depends on the negative effects induced on the original photo. As a consequence,
79 IQ could be estimated as the similarity between the transformed image and the original image. In
80 other words, IQ is the measurement of how close the transformed versions and the original one are. In
81 former way, the simplest metric based on signal processing theory is to compute the differences of
82 corresponding pixels between versions. Classical IQA methods are based on distortions computed on
83 pixel values such as mean squared error, root mean squared error, peak signal to noise ratio. Although
84 it is a simple approach, obtained results are not really suitable for IQA because human perception is
85 more complicated than the way machines process signals. The computed difference does not always
86 match with visual perception. A more efficient approach is based on human visual system properties.
87 That approach is based on psychological and physical characteristics of human visual system to
88 compute visual quality of photos such as luminance, contrast, structure, fidelity criterion, similarity
89 indexes [2–8].

90 IQ is not always estimated through the similarities between the transformed version and the
91 original version because original photos without any modifications are not always the best versions.
92 Although original images are not affected negatively by processing methods, artifacts and distortions
93 could still exist in those photos because of many factors such as camera sensors, camera settings,
94 brightness conditions, motions. In the cases of image restoration and image enhancement processes,
95 IQ of some transformed versions could be better than that of the original one so IQ in this case is based
96 on the enhancement of bad visibility [9]. IQ is also based on viewers' preference [10]. In some cases,

97 an appealing version could be preferred to a version that is more similar to the reference than the
98 appealing one. It appears that the most preferred version is not necessarily the closest to the reference
99 image.

100 On the side of NRIQA, IQ is on the contrary based on viewers' background including preference,
101 visual aesthetic perception, color memory, naturalness perception. A lot of metrics have been presented
102 and validated on results coming from subjective experiments to estimate IQ in this context [11–23].
103 Ke et al. [11] define IQ based on abstract aesthetic aspects such as composition, color, lighting to classify
104 professional and snapshot photos. Additionally, they consider simplicity, realism and photography
105 technique as the three main factors producing a high quality image. Similarly, in [12,14,16] IQ is defined
106 from the perspective of visual aesthetic. According to photography rules, Tang et al. [16] propose an
107 NRIQA metric based on professionals' views including composition, lighting, color arrangement,
108 camera settings and topic emphasis. In another approach, Hosu et al. [21] define IQ as a technical
109 concept related to some types of degradations such as over-saturation, noise, aliasing, motion blur,
110 wrong exposure, over-sharpening, color fringing, and JPEG artifacts. Using the same approach, IQ
111 is based on visual distortions induced by technical causes (noise, blur, JPEG compression) in [15].
112 Beside that, IQ of tone-mapped images is defined in a different way. Yaacoub et al. [18] consider tone
113 mapped IQ as the balance between luminance contrast and naturalness. In [17], it is assumed that
114 a high quality tone-mapped image maintains global information, details and naturalness so IQ in
115 this case is described as the combination of luminance, structure and naturalness. In another study,
116 Jiang et al. [19] assume that IQ of tone-mapped images could be affected because of underexposure,
117 overexposure and losses in Image Naturalness (IN) and Image Aesthetic (IA). They describe IQ by
118 using three factors including brightness in the brightest and darkest regions, IN and IA. Using a
119 different approach, Leisti et al. [13] defined IQ based on low level attributes related to physical aspects
120 (sharpness, lightness, graininess) and high level attributes related to abstract aspects (individual
121 feelings, viewers' experiences, naturalness). In [23], Varga et al. pointed out that first digit distributions
122 based metrics of high quality images, extracted from multiple domains match well with Benford's
123 law. That study demonstrated that first digit distribution are quality-aware features and it is possible
124 to archive high performance in NRIQA with those features. Based on a different approach, multiple
125 global average pooling architecture is used for IQA in [22]. Specifically, a deep model containing
126 multiple inception blocks attached with average global pooling layers to extract features. Instead of
127 taking patches from the input image the whole image is passed through a pre-trained model so the
128 proposed architecture could work with images in various resolutions. Based on the obtained results,
129 that work might be a potential base to improve the performance of IQA.

130 Another IQA protocol mentioned here is reduced reference IQA. Partial information only is
131 provided about the reference image. Characteristics or features such as histogram, color saliency map,
132 sharpness map are extracted. In reduced reference IQA, reduced references somehow are like human
133 memory. A person sometimes feels that he/she has seen the scene of a photo but without remembering
134 all details. Several IQA metrics based on reduced reference have been introduced: in [24], the IQA
135 metric is based on a divisive normalization image representation coming from a Gaussian scale mixture
136 statistical model of image wavelet coefficients, in [25], the IQA metrics is based on a linear relationship
137 between full reference and reduced reference structural similarity index measures, in [26], the IQA
138 metric is based on an orientation selectivity mechanism for visual content extraction, in [27], the IQA
139 method is based on saliency maps and texture similarity between high-resolution and low-resolution
140 photos.

141 Generally, in either FRIQA or other IQA protocols, IQ refers to the measurement of how photos
142 satisfy viewers. The satisfaction of viewers mainly depends on visual aesthetic perception and the
143 feeling of how close photos and real scenes are. Therefore, it could be seen that the two factors: IA and
144 IN play important roles in IQ.

145 2.2. Image Aesthetic

146 The questions of how a photo is captured as well as how a viewer enjoys and criticizes the photo
147 lead to the visual perception of aesthetics in photography so a part of IA perception is based on
148 photographic rules. Many studies have been made to modelize IA. In [28], IA is described based on
149 rules of composition, depth of field, salient object, opposing colors and natural illumination. In another
150 study, Marchesottian et al. [29] introduce a description of IA using Bag of visual words descriptors,
151 Fisher vector and GIST descriptors. Beside that, in [30], IA is evaluated based on the combination of
152 simplicity, patterns of harmony and rhythm, colorfulness, composition and sharpness. In a similar
153 approach, Aydin et al. [31] introduce an IA signature concept based on sharpness, clarity, tone, depth
154 and colorfulness features. However, not all aesthetic aspects are describable so it does not mean that
155 following photography rules always produces a high aesthetic photo and on the contrary, a beautiful
156 photo might not follow those rules. As a consequence, there is a part of subjectivity in IA perception.
157 Deep learning approach might be a good solution to describe the subjective aspects. Various researches
158 of IA using deep learning have been proposed such as the Image Aesthetic Assessment (IAA) model
159 based on the combination of a retrieval system and a deep Convolutional Neural Network (CNN)
160 presented in [32], the double-column deep CNN architecture using two parallel CNNs based on global
161 and local features proposed in [33], the CNN including 3 kinds of layers: transferred layers, scene
162 convolutional layers and fully connected layers evaluating IA of multi-scenes [34], the IAA model
163 based on deep learning technique, image classification and image segmentation introduced in [35].
164 Besides, Hii et al. [20] proposed a deep model exploiting multiple inception blocks pooled by global
165 average pooling. They also integrated textual information with visual information to perform IQA.
166 Experimental results in that work demonstrated a good performance of the proposed architecture.

167 2.3. Image Naturalness

168 Different definitions of IN have been introduced. In [36–38], IN is described as the degree of
169 correspondence between a photo displayed on a device and the memory about the real-life scene.
170 In [39], Cadik et al. introduce IN as the degree of correspondence between a scene (seen directly)
171 and the corresponding scene in photos based on some technical criteria: brightness, contrast, colour
172 reproduction, reproduction of details, simulation of glare, visual acuity and artifacts. In another study,
173 Jiang et al. [19] define IN based on the differences between photos taken with normal exposure and
174 abnormal exposure in which unnatural photos are described as over or under exposure photos and
175 natural photos are considered as photos captured with normal exposure. In some researches, IN
176 features have been employed for IQA [17–19]. Besides, in [1], IN is based on artifacts induced by some
177 image processing methods (such as halos, blur, lost details) and on the individual feeling (memory,
178 opinion, background).

179 It could be seen that many efforts of simulating human visual perception have been made.
180 Machines have been trained to understand and measure IQ like humans. In order to understand
181 more about IQ, IA and IN to simulate human vision perception, the correlations between the two
182 aspects of IQ: IA and IN are studied in this work. Obviously, IA and IN have been described in various
183 computational ways in previous researches as presented above so a computational approach might be
184 an acceptable choice to study IA and IN.

185 3. Potential Relationships between IA and IN

186 In order to answer the main purpose of this section, three sub tasks are considered. The first one
187 is to estimate the correlation between IA features and IN features. Secondly, the influence of IA and
188 IN features on Image Naturalness Assessment (INA) and IAA respectively is measured. The last task
189 is about the equivalence between high aesthetics and naturalness on the one side and between low
190 aesthetics and unnaturalness on the other side.

191 Although many IA datasets and IN datasets have been introduced for decades, a dataset with
192 both IA and IN ground-truth from subjective experiments does not exist. Therefore, instead of using a
193 dataset with both subjective IA and IN ground-truth to evaluate the relationships between IA and IN,
194 an IA dataset and an IN dataset with subjective ground-truth are used and the IN of the IA dataset
195 and the IA of the IN dataset are computed by an INA model and an IAA model respectively.

196 In this research, an IA dataset coming from [35] and an IN dataset coming from [1] are considered.
197 On one side, the IA dataset contains 1200 high aesthetic images and 1200 low aesthetic images coming
198 from the CUHKPQ dataset [16]. Each photo of the CUHKPQ dataset is evaluated by ten observers and
199 a photo is considered as “high aesthetic” if at least eight of the ten viewers consider its aesthetic as
200 “high” [16]. Similarly, a photo is labeled as “low aesthetic” if at least eight of the ten viewers consider
201 its aesthetic as “low”. On the other side, the IN dataset contains 355 natural photos and 515 unnatural
202 photos assessed each by nine observers. A photo is labeled as “natural” if at least eight of the nine
203 viewers consider it as “natural” and similarly, a photo is labeled as “unnatural” if it is assessed as
204 “unnatural” by at least eight of the nine observers [1].

205 Beside that, an IAA model based on the IA feature set learned from [35] and an INA model based
206 on the IN feature set coming from [1] are considered in this paper because of their good performance.
207 Firstly, the IA feature set contains 1024 global features (features learned from the whole image) learned
208 by a deep CNN having a typical architecture with an input layer, an output layer and five convolutional
209 blocks. Specifically, Two convolutional layers and a pooling layer are placed in each block. There
210 are 64×2 , 128×2 , 256×2 , 512×2 , 1024×2 kernels in the five blocks respectively (two layers in each block).
211 The last layer contains two outputs corresponding to the two categories: high visual aesthetic image
212 and low visual aesthetic image. The model is trained on over 18,000 high and low visual aesthetic
213 photos coming from the CUHKPQ dataset [16] and the obtained accuracy is quite impressive at
214 0.914. Secondly, IN features are studied in [1]. Various models including Xception [40], NASNet [41],
215 MobileNet [42], InceptionNet [43], VGG16 [44] and ResNet [45] pre-trained on over 14 million images
216 of the ImageNet dataset for the task of image classification were considered in that work. The feature
217 selection process described in [46] is applied to select the most relevant features to the task of INA
218 from features of each pre-trained model. Those models are then transferred to the new purpose of
219 INA by replacing the last layers (all the fully connected layers) of those models and re-training those
220 layers on the considered IN dataset for the INA task. After training and testing the models using the 9
221 reduced feature sets, the highest overall accuracy (0.865) and the best loss (0.139) are obtained with
222 the model using the features learned from the ResNet extractor. In this case, there is no re-trained
223 ResNet layers. The model without the last layer (the fully connected layer) is considered as the feature
224 extractor for the proposed model. Specifically, 425 learned features are selected from the 2048 ResNet
225 features by applying the Relief based feature reduction algorithm [46]. The overall accuracy of the
226 classification is quite good at 0.865.

227 3.1. IA and IN Feature Correlation Analysis

228 In this work, two common correlation coefficients are employed to measure the correlation
229 between IA and IN features. The first one is the Pearson correlation coefficient [47]. It is a measure of
230 linear correlation between two sets of data. It is computed as the ratio between the covariance of two
231 variables and the product of their standard deviations. In other words, the Pearson correlation draws
232 a line of best fit through the data of two variables and calculates the effect of change in one variable
233 when the other variable changes. For example: the positive correlation between child’s age and height,
234 (in most cases, a child’s height will keep increasing as his/her age increases), the negative correlation
235 between vehicle’s speed and traveling time (if a vehicle increases its speed, the time it takes to move
236 decreases, and vice versa). The second coefficient is the Spearman rank correlation [48]. The Spearman
237 rank correlation between two variables is equal to the Pearson correlation between the rank values
238 of those two variables. Both coefficients range from -1 to 1 in which values near 1 and -1 refer to a
239 perfect correlation in which 1 means that if the value of one variable increases, the value of the other

240 variable increases too while -1 means that if the value of one variable increases, the value of the other
 241 variable decreases. Value 0 reflects no correlation. For the sake of simplicity, the correlation absolute
 242 values only are considered. They range from 0 to 1 and the higher absolute value refers to the higher
 243 correlation.

244 To measure the correlation between the IA feature set and the IN feature set, each feature of
 245 the two feature sets is computed on n images to form a n -dimensional vector. The correlation $cor[i, j]$
 246 between the i -th IA feature and the j -th IN feature is computed as the absolute value of the correlation
 247 between the two corresponding n -dimensional vectors. The most correlated IN feature to the i -th IA
 248 feature is determined as the IN feature having the highest correlation ($maxCor_i^{IA}$) to the IA feature as
 249 in (1) in which 425 is the number of IN features. Similarly, the most correlated IA feature to the j -th IN
 250 feature is determined as the IA feature having the highest correlation ($maxCor_j^{IN}$) to the IN feature as
 251 in (2) in which 1024 is the number IA features. Histograms with ten bins are built based on the highest
 252 correlation of each feature computed on the IA dataset [35] and the IN dataset [1] as in Figure 4 and 5.

$$maxCor_i^{IA} = \max_{j=1}^{425}(cor[i, j]) \quad (1)$$

$$maxCor_j^{IN} = \max_{i=1}^{1024}(cor[i, j]) \quad (2)$$

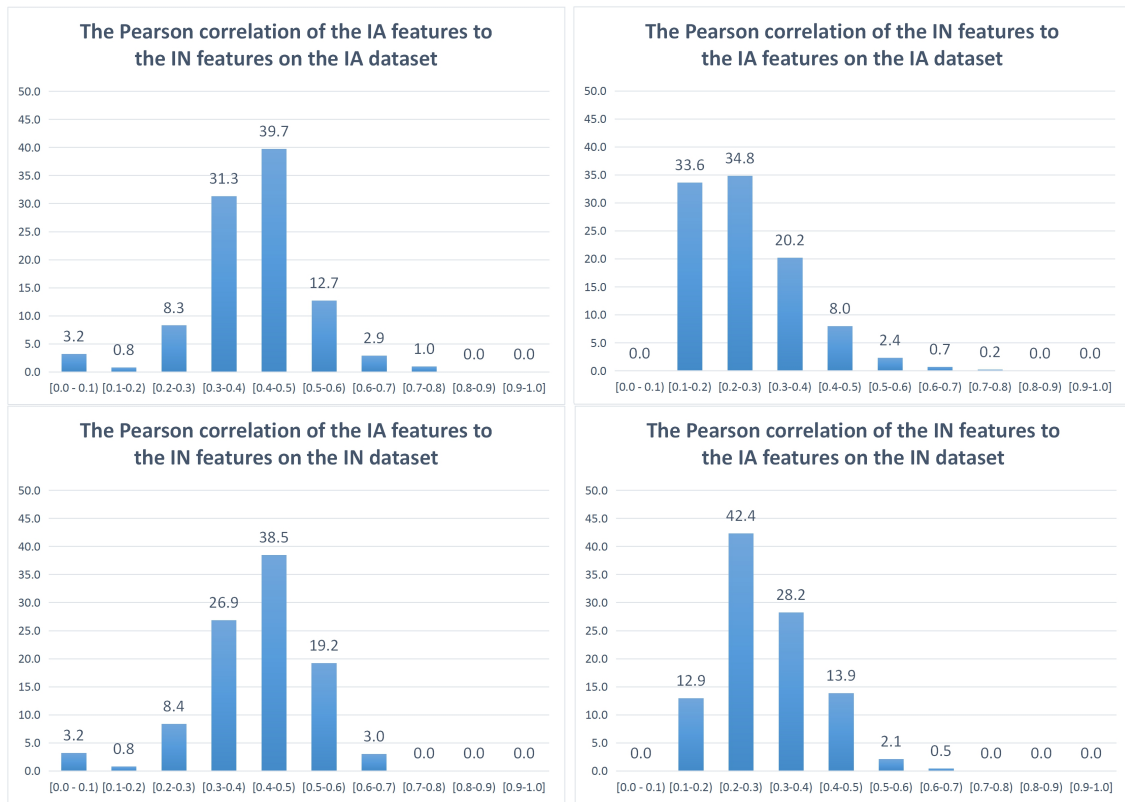


Figure 4. Pearson correlation between IA features and IN features.

253

254 Figure 4 shows the Pearson correlation between IA features and IN features. Specifically, the
 255 horizon axis represents groups of correlation values while the vertical axis shows the proportion of
 256 features (in percentages). On the IA dataset, it appears that a small part of IA features are highly
 257 correlated (the highest correlation of the features is higher than 0.5) to the IN features (15.6%) while a
 258 majority of IA features is moderately correlated (the highest correlation of the features ranges from 0.3

259 to 0.5) to the IN features (71%). Besides, a minority of IA features (13.4%) is weakly correlated (the
 260 highest correlation of the features is lower than 0.3) to the IN features. There is a similar trend on the
 261 IN dataset where there are 22.2%, 65.4% and 12.4% of IA features are highly correlated, moderately
 262 correlated and weakly correlated to the IN features respectively. In contrast, the IN features seem to be
 263 less correlated to the IA features since a majority of IN features is weakly correlated to the IA features
 264 (68.4% on the IA dataset and 55.3% on the IN dataset) while only 3.3% and 2.6% of IN features are
 highly correlated to the IA features on the IA and IN datasets respectively.

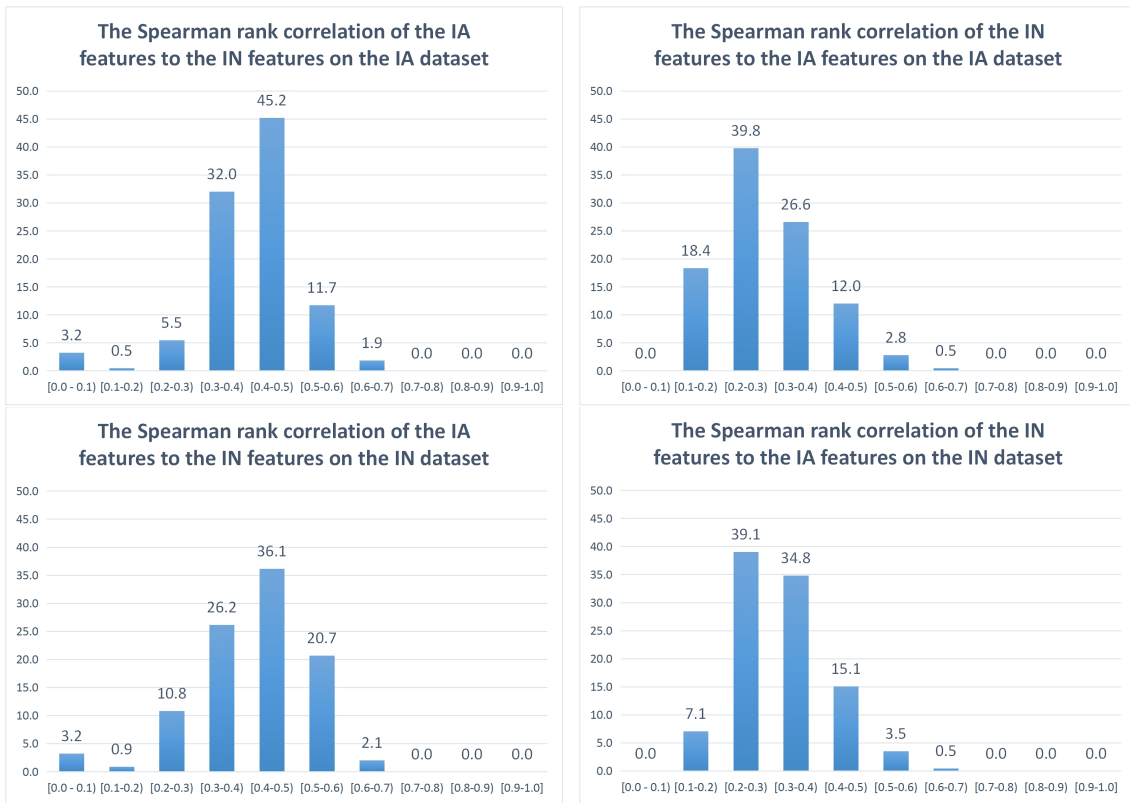


Figure 5. Spearman correlation between IA features and IN features.

265 Similarly, Figure 5 shows the Spearman rank correlation between IA features and IN features.
 266 The results based on the Spearman rank correlation coefficient are quite similar to the results with the
 267 Pearson correlation coefficient. A majority of IA features is highly correlated and moderately correlated
 268 to the IN features (13.6%, 77.2% on the IA dataset and 22.8%, 62.3% on the IN dataset respectively)
 269 while a significant part of IN features are weakly correlated to the IA features (58.2% and 46.2% on the
 270 IA dataset and the IN dataset respectively). Beside this, only 3.3% and 4% of IN features are highly
 271 correlated to the IA features on the IA dataset and the IN dataset respectively.
 272

273 Generally, although there are few differences between the results based on the Pearson correlation
 274 and based on the Spearman rank correlation, both results refer to the same general conclusion that IA
 275 features have a quite significant correlation with IN features but the correlation between IN features
 276 and IA features is much weaker.

277 3.2. Are IN and IA Independent or Dependent?

278 Although the correlation between IA and IN features has been estimated in the previous section,
 279 the meaning of the correlation between IA and IN has not been clarified. The considered IA and IN
 280 features are learned automatically by deep CNNs [1,35] so they are abstract and not easy to understand.
 281 In this subsection, the idea is to check if IN and IA are independent so that they describe two different

282 aspects of IQ or not. To do so, we are going to study if there is an overlap between IA features and IN
283 features first and then the influence of IN features on IAA and the influence of IA features on INA are
284 going to be evaluated.

285 3.2.1. Influence of IN Features on IAA

286 According to the results of section 3.1, the correlation of IN features with IA features is low. Based
287 on Pearson coefficient, only 14 IN features (3.3%) are highly correlated to IA features while the number
288 of weakly correlated IN features is 291 (68.4%). Beside that, there are 120 IN features moderately
289 correlated to IA features (28.2%). It appears that a majority of IN features is not overlapping with IA
290 features but the number of moderately correlated IN features is significant (28.2%).

291 In order to evaluate the influence of IN features on IAA, we propose to train an IAA model by
292 considering IN features only including 14 highly correlated, 120 moderately correlated and 291 weakly
293 correlated features. The performance of this model is compared with the one based on IA features
294 only and the one based on the combination of IA and IN features. Figure 6 presents the process of the
295 experiment and the results. The proposed IAA model containing an input layer (the number of input
296 nodes is the number of input features) and an output layer (one output node with sigmoid activation
297 function, a very simple linear model is considered because in this study, we want to focus on the
298 impact of features instead of the architecture of the model) is trained to perform a binary classification
299 between high and low aesthetic photos. The Adam optimizer and a binary cross-entropy loss function
300 are used and the batch size is assigned to 100. The learning rate and the number of iterations are set to
301 5×10^{-4} and 150 respectively. The IAA model is trained and tested on the IA dataset coming from [35].
302 The IA dataset labeled by humans is split into a training set containing 1600 images (two thirds of the
303 dataset) and a testing set including 800 images (one third of the dataset).

304 Looking at the results in Figure 6, although the performance of the IAA based on IN features is
305 lower than the one based on IA features (0.915 ± 0.019 versus 0.930 ± 0.018), it is quite impressive.
306 The moderately correlated IN features could be the reason for the good performance with IN features
307 since the IAA based on them has a good performance (0.911 ± 0.020). Besides, the number of highly
308 correlated features being small (14 features), the performance of the IAA based on them is bad
309 (0.769 ± 0.029). The results with the weakly correlated IN features are not very impressive since
310 the IAA based on them has a lower performance at 0.874 ± 0.023 . Although those features are not
311 overlapping with IA features, they are not related to the IAA task. It could explain the slight increase
312 of accuracy from 0.930 ± 0.018 to 0.946 ± 0.016 when considering the IAA based on the combination of
313 IA features and IN features.

314 3.2.2. Influence of IA Features on INA

315 Section 3.1 shows a significant correlation between IA features and IN features. Based on Pearson
316 coefficient, 228 (22.2%) and 669 (65.4%) IA features are respectively highly correlated and moderately
317 correlated to IN features while the number of weakly correlated IA features is 127 (12.4%). It appears
318 that there is a significant part of IA features has overlaps with IN features.

319 Similarly, INA based on IA features only is investigated in which the IA features include the
320 228 highly correlated, the 669 moderately correlated and the 127 weakly correlated features. The
321 performance of INA based on IA features is compared with the one based on IN features only and
322 the one based on the combination of IA features and IN features. The process of the experiment and
323 experiment results are presented in Figure 7. Starting with the considered IN dataset [1], in order to
324 balance the number of natural and unnatural photos in the training set, a data augmentation process
325 is applied to generate augmented versions of natural and unnatural photos by re-scaling, cropping
326 and padding. Naturalness labels of augmented versions are kept the same as the labels of the original
327 images. A training set (generated from 84% of the dataset) containing 1704 natural photos (284 original
328 photos \times 6 data augmented versions), 1776 unnatural photos (444 original photos \times 4 data augmented
329 versions) and a testing set containing 71 natural photos and 71 unnatural photos (16% of the dataset

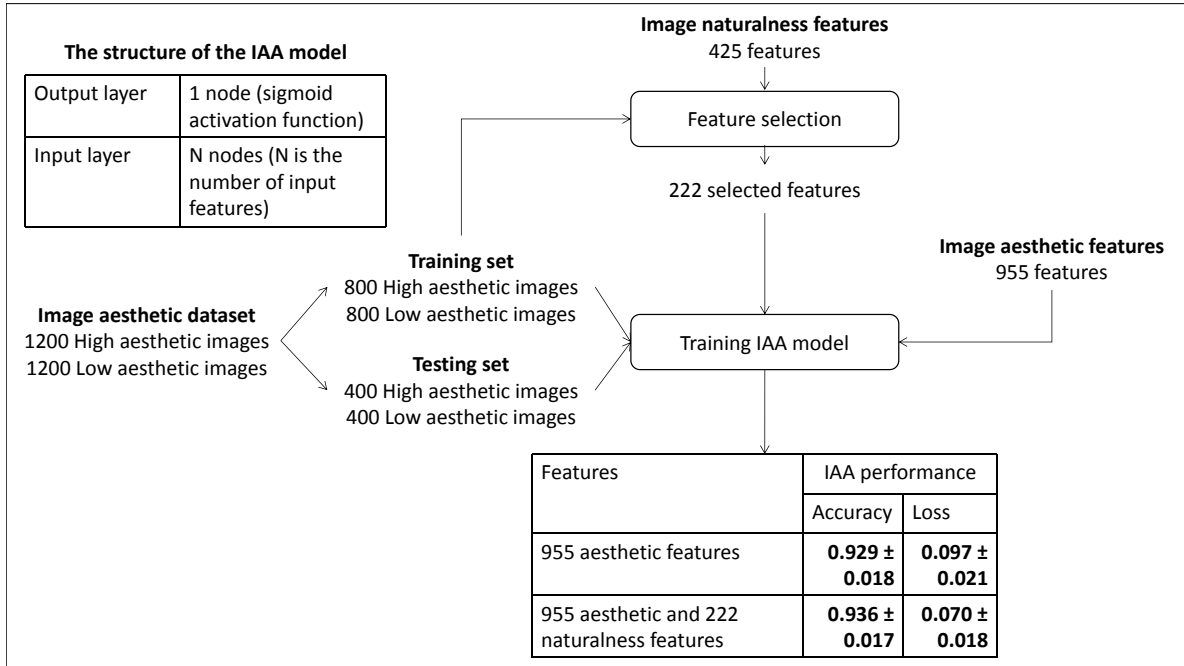


Figure 6. Process of the experiment evaluating how IN features improve the performance of IAA.

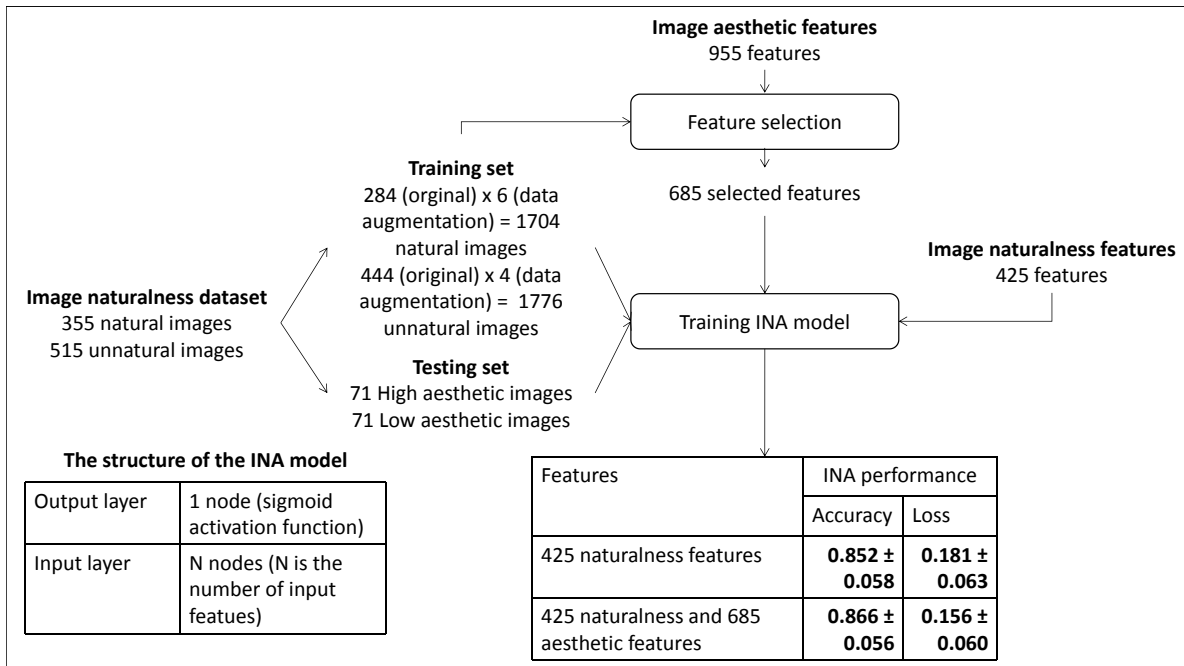


Figure 7. Process of the experiment evaluating how IA features improve the performance of INA.

without data augmentation) are extracted. The structure of the INA model and training parameters are set the same as in the previous experiments.

The experiment results in Figure 7 show that although the performance of the INA based on IA features is lower than the one based on IN features (0.835 ± 0.061 versus 0.852 ± 0.058), that result is quite good. The highly correlated and moderately correlated IA features could be the reason for the good performance of the INA based on IA features since the INA based on them has good results (0.814 ± 0.064 and 0.822 ± 0.063 respectively). Besides, the results with the weakly correlated IA features are not good since the INA based on them has a much lower performance at 0.773 ± 0.069 . Experiment results reflect that IA features do not help improving the INA performance significantly since the accuracy of INA based on IN features only is 0.852 ± 0.058 while this value of INA based on the combination of IN and IA features just increases insignificantly to 0.880 ± 0.053 .

The obtained results prove that there is an overlap between IA and IN features and explain why the performance of IAA based on IN features only and the performance of INA based on IA features only are quite good. However, there are uncorrelated parts between IA and IN features. The performance of IAA and INA based on the uncorrelated features is not really good so those IN and IA features might not be related to IAA and INA tasks respectively. As a consequence, the combination of correlated and uncorrelated features do not help improving significantly the performance of IAA and INA.

3.3. Relation between Naturalness / Unnaturalness and Low / High Aesthetics

3.3.1. Are Natural Images High Aesthetic Ones and Unnatural Images Low Aesthetic Ones?

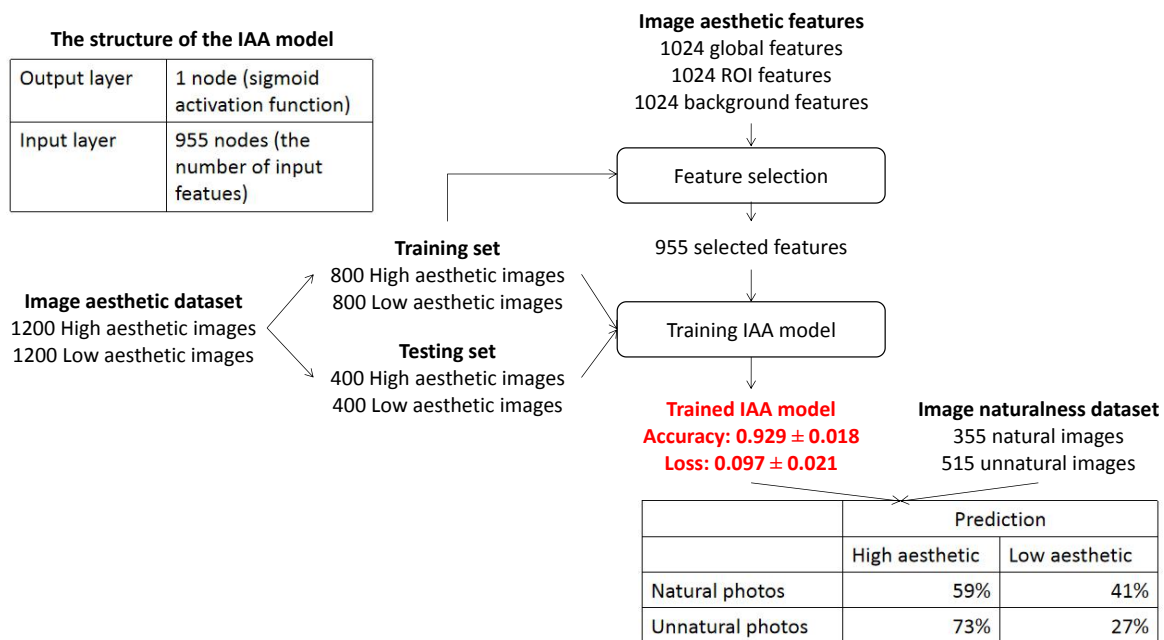
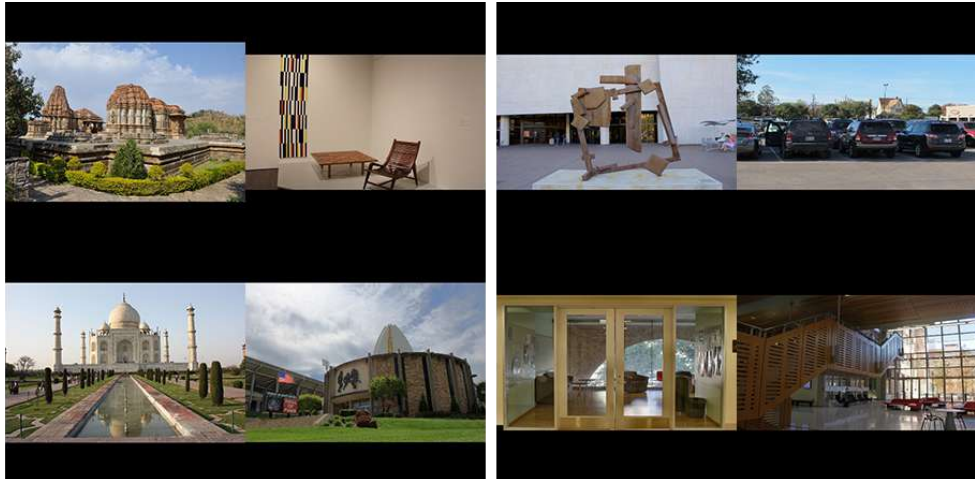


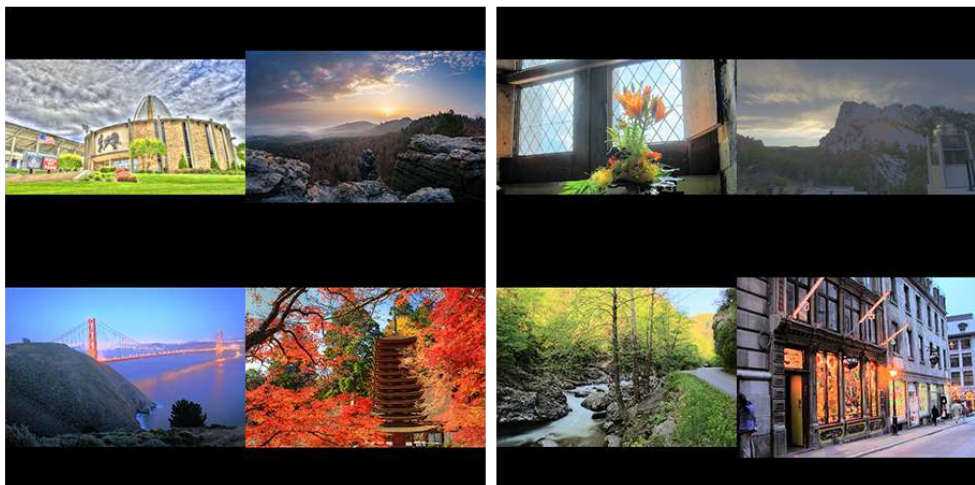
Figure 8. Process of the experiment evaluating IA of the IN dataset images.

To answer that question, IA of natural and unnatural images is investigated. Figure 8 shows the proposed process of evaluating IA of the two image categories. In part 3.2.1, the model based on the combination of IA features and IN features has the highest performance (0.946 ± 0.016) so it is used to make the distinction between high aesthetic images and low aesthetic images. As a consequence, this model is used to assess the IA of natural and unnatural photos of the IN dataset [1].

Natural photos



Unnatural photos



Predicted as high aesthetic

Predicted as low aesthetic

Figure 9. IAA samples of natural and unnatural images.

355 According to experiment results, 28% of the natural images are predicted as high aesthetic and
 356 61% of unnatural images are predicted as low aesthetic. On the contrary, a significant part of natural
 357 photos (72%) is assessed as low aesthetic and an insignificant part of unnatural photos (39%) is
 358 predicted as high aesthetic. So there is clearly not a cause to effect relation between naturalness and
 359 high aesthetics and unnaturalness and low aesthetics. Natural photos are classified as low aesthetic
 360 more than unnatural ones. The reason could be the lack of post processing in natural photos making
 361 those photos boring.

362 3.3.2. Are High Aesthetic Images Natural Ones and Low Aesthetic Images Unnatural Ones?

363 Similarly, in order to answer the question of this part, the IN of high and low aesthetic photos
 364 is investigated. The model learned to assess IN is presented in Figure 10. According to part 3.2.2,
 365 the INA model based on the combination of IA features and IN features has the best performance
 366 (0.880 ± 0.053) so it is used to classify natural and unnatural in this part. The IN of high aesthetic
 photos and low aesthetic photos of the IA dataset [35] is predicted by the INA model. According to

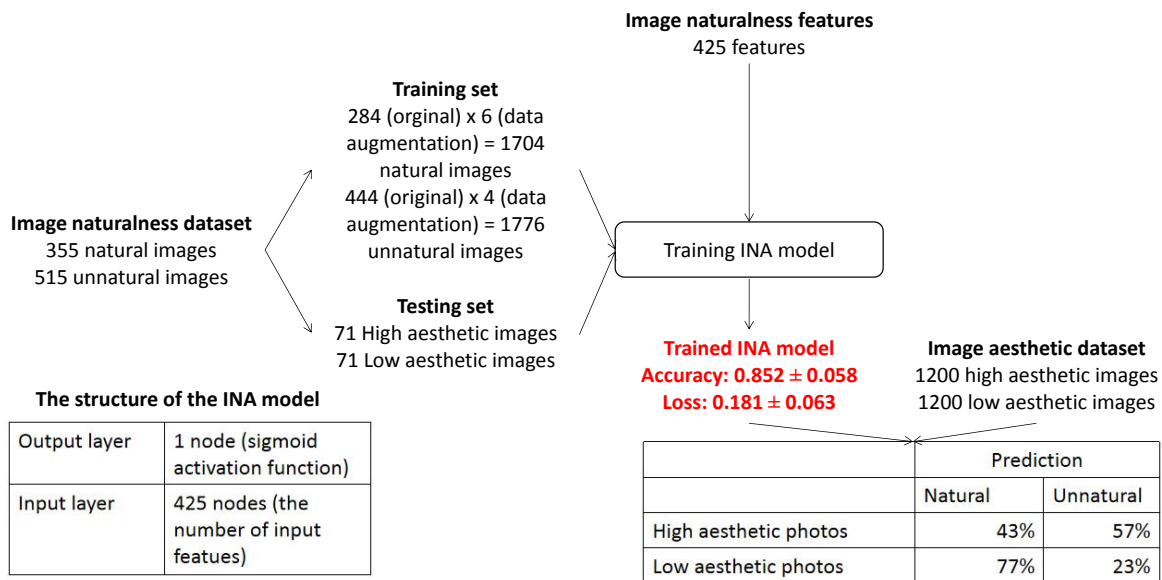


Figure 10. Process of the experiment evaluating IN of photos in the IA dataset.

367 the experiment results, 35% of the high aesthetic photos are predicted as unnatural while the majority
 368 of low aesthetic photos (89%) that are mostly not post-processed is assessed as natural. It appears
 369 that a high aesthetic photo does not mean a natural photo and a low aesthetic photo is not always
 370 unnatural.
 371

372 3.3.3. IA and IN score correlation

373 Additionally, the IA scores and the IN scores predicted by the IAA and the INA models on photos
 374 are considered as two vectors. The Pearson correlation and the Spearman rank correlation between
 375 the two vectors are computed and the results are presented in Table 1. It appears that there is a weak
 376 negative correlation between the IA scores and the IN scores even on natural images, unnatural images,
 377 high aesthetic images, low aesthetic images or all images.

378 Experiment results prove that there is no direct correlation between IA and IN since natural
 379 / unnatural photos are not always considered as high / low aesthetic respectively and vice
 380 versa. Samples of IAA and INA are presented in Figure 9 and Figure 11 respectively. Obviously,

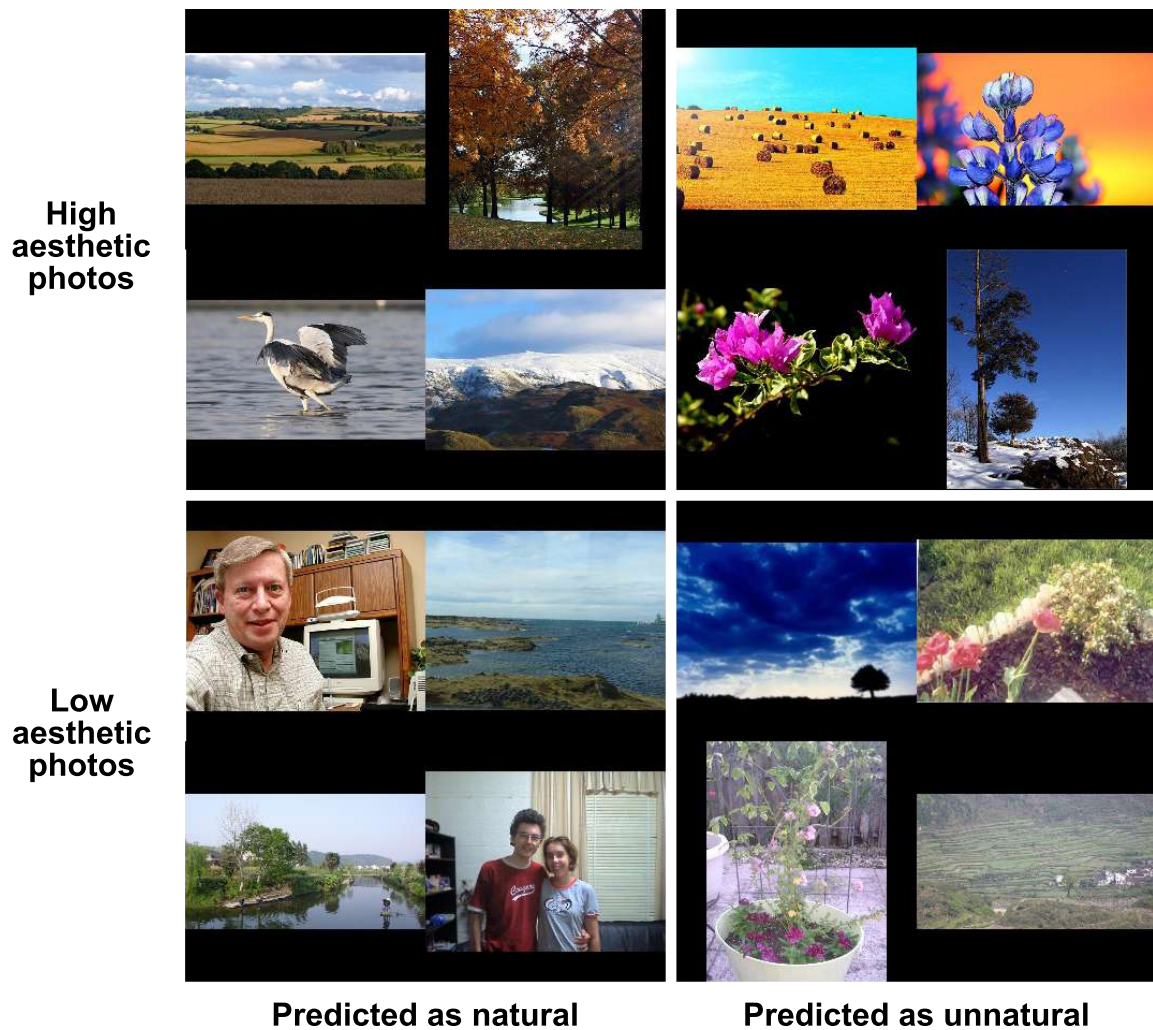


Figure 11. INA samples of high and low aesthetic photos.

Table 1. The correlation between IA scores and IN scores.

Correlation between IA and IN scores computed on	Pearson correlation	Spearman rank correlation
Natural images	-0.078	-0.105
Unnatural images	-0.129	-0.066
High aesthetic images	-0.094	-0.145
Low aesthetic images	-0.139	-0.087
All images	-0.191	-0.218

381 abusing enhancement methods that increase perceived aesthetic quality could provoke artifacts from
382 imperceptible to obvious (over-enhancement) so the increase of IA could lead to the decrease of IN
383 (see unnatural photos predicted as high aesthetic in Figure 9 and high aesthetic photos predicted as
384 unnatural in Figure 11) or even to the decrease of IQ generally. On the contrary, when comparing
385 a photo re-produced by an adjustment method such as a tone-mapping operator and other single
386 exposure versions, a tone mapped photo could be more natural than a single exposure version of
387 the same scene with deep, lively and realistic colors and contrast (see natural photos predicted as
388 high aesthetic in Figure 9). Adjusted photos could be more appealing and interesting because of the
389 uniqueness (compared with normal single exposure images that cannot preserve high contrast, deep
390 colors of the real scenes). However, when a photo is too faithful and familiar to observers so they
391 might not be interested in the photo (see low aesthetic photos predicted as natural in Figure 11).

392 Generally, although IN and IA might have some correlations, they are nevertheless two different
393 notions referring to two different aspects of IQ.

394 4. How do IA and IN Affect Viewers' IQ Perception in Different Contexts?

395 IQ as well as IA and IN are both related to quality of service [49] and quality of experience [50,51].
396 Quality of experience is based on human factors (individual properties, attitudes: visual and auditory
397 acuity, gender, age, cognitive processes, socio-cultural and economic background, expectations) and
398 contextual factors (experiment conditions: lightness, reference, distance, time, location). In contrast,
399 quality of service refers to photos' properties and characteristics. The main question to be investigated
400 is "is IQA related to quality of experience or quality of service?". Some people advocate that there
401 are clear features, properties of photos deciding IQ while opponents suggest that IQ is driven by
402 individual opinions, experiences and context. The answer could be somewhere in the middle since
403 individual feelings and experiment context have great influences on IQ perception while intrinsic
404 properties of photos can also affect IQ.

405 The main goal of this section is to study the relationships between IQ and quality of experience.
406 Specifically, the influence of IA and IN on IQ in different contexts is investigated. The influence of IA
407 and IN in two cases FRIQA and NRIQA is considered. In order to clarify the relations between IA, IN
408 and IQ and how the influence of IA and IN on IQ is affected by experimental contexts, the idea is to
409 investigate the correlation between IA and IQ, and the one between IN and IQ in two cases: with and
410 without reference.

Figure 12 illustrates the whole process of estimating the correlation between IQ and IA and the one between IQ and IN. Firstly, starting with the subjective IQA experiment presented in [10], 10 HDR scenes are considered and nine different tone-mapped versions are generated from each HDR scene by using simple linear clipping with inverse gamma correction, Drago [52], iCAM06 [53], Mantiuk [54] and Mai [55] tone-mapping operators with different parameters settings. Those photos are categorized by scene so there are nine tone-mapped versions and an HDR photo in each category. Possible tone-mapped image pairs in each category have been showed to 20 naive observers with and without reference. In their first test, observers had to choose which image they prefer between the two mapped versions of a given pair. Those images have been displayed on an SDR screen with a background luminance of 50 cd/m^2 and a max luminance of 200 cd/m^2 . In the second test, not only a given pair but also a reference displayed on an HDR screen (background luminance of 100 cd/m^2 and max luminance of 4000 cd/m^2) have been showed to viewers each time and they had to answer the same question as in the first test (see Figure 13). The experiments were performed according to the ITU-R BT.500-11 for a subjective experiment. In those experiments, some images closer to the reference are preferred. However, some images less similar to the reference are sometimes preferred because they are more appealing. This subjective experiment is based on viewers' global preference only: IA and IN are not mentioned in that experiment. Considering an HDR scene, Figure 14(a) shows the pair comparison matrix of the nine versions mapped from the HDR scene in which $PCM[i, j]$ presents the number of time the i^{th} tone-mapped version is preferred when comparing it to the j^{th} version. It

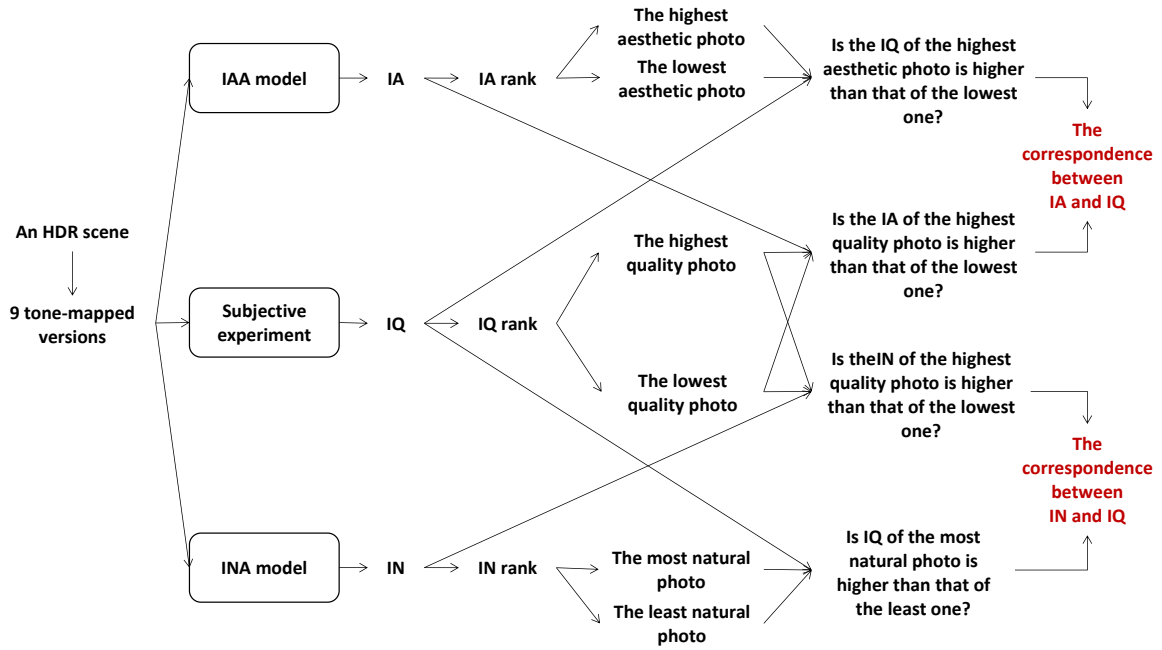


Figure 12. Process of estimating the correlation between IQ and IA and the one between IQ and IN.

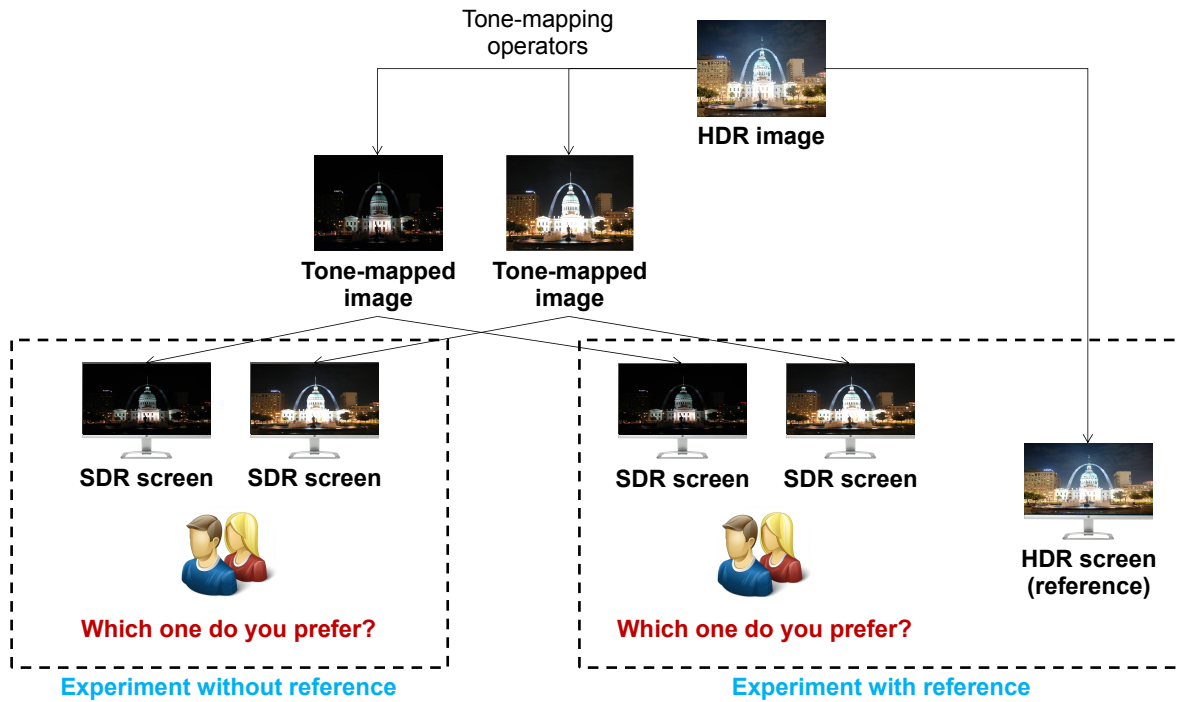


Figure 13. IQA experiment based on human preference with and without reference.

appears that the total number of observers for each pair is not the same so the Bradley-Terry score Matrix is computed as in (3). IQ of the i^{th} tone-mapped version is then computed as in (4). The nine tone-mapped versions of each HDR scene are ranked based on IQ values. Figure 14(b) and 14(c) show Bradley-Terry score Matrix, IQ values and the IQ rank of the considered photos.

$$BTM[i, j] = \frac{PCM[i, j]}{PCM[i, j] + PCM[j, i]} \quad (3)$$

$$IQ[i] = \frac{1}{9} \sum_{j=1}^9 BTM[i, j] \quad (4)$$

0	5	2	1	6	4	15	5	11
3	0	5	3	9	6	9	7	4
6	11	0	8	3	7	8	5	11
9	4	4	0	11	3	7	7	12
4	5	1	2	0	6	8	4	8
6	6	5	5	4	0	8	6	2
0	0	0	0	0	2	0	3	2
3	2	2	3	7	5	10	0	9
0	1	2	1	2	5	8	2	0

(a)

0.00	0.63	0.25	0.10	0.60	0.40	1.00	0.63	1.00
0.38	0.00	0.31	0.43	0.64	0.50	1.00	0.78	0.80
0.75	0.69	0.00	0.67	0.75	0.58	1.00	0.71	0.85
0.90	0.57	0.33	0.00	0.85	0.38	1.00	0.70	0.92
0.40	0.36	0.25	0.15	0.00	0.60	1.00	0.36	0.80
0.60	0.50	0.42	0.63	0.40	0.00	0.80	0.55	0.29
0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.23	0.20
0.38	0.22	0.29	0.30	0.64	0.45	0.77	0.00	0.82
0.00	0.20	0.15	0.08	0.20	0.71	0.80	0.18	0.00

(b)

0.51	4th
0.54	3rd
0.67	1st
0.63	2nd
0.44	6th
0.46	5th
0.07	9th
0.43	7th
0.26	8th

(c)

Figure 14. Process of pair comparison matrix analysis: (a) The pair comparison matrix of an HDR scene in which $PCM[i, j]$ is the number of time I_i is preferred when comparing I_i to I_j , (b) Bradley-Terry score Matrix of each HDR scene (c) IQ and ranks of tone-mapped versions of each HDR scene.

411

412

413

414

415

416

417

418

419

420

421

422

There are no subjective results of IA and IN for those tone-mapped images. Therefore, in the second step (cf. first line and third line of Figure 12), IAA and INA models used in part 3.3 are applied on the tone-mapped versions of each HDR scene in order to predict their IA and IN scores respectively. Firstly, the perception of IA is general, viewers focus on aesthetic criteria (how does the image satisfy their expectations) instead of technical criteria (how is the image generated). In addition to this, The IAA model was trained on a general image set containing different types of images (single exposure images, post-processed images, tone mapped images) so it refers to general IA. Secondly, the INA model was trained on a dataset containing different types of images in which over 50% of the dataset are tone-mapped images so it is able to predict IN of tone-mapped images. That is the reason why the IAA model has been used on tone-mapped images in this case to evaluate the impact of IA and IN on IQ perception.

423

424

425

426

427

428

429

430

431

432

433

434

435

Output scores respectively range from 0 to 1 referring to IA ([0, 0.5) means low aesthetic while [0.5, 1] means high aesthetic) and IN ([0, 0.5) means unnatural and [0.5, 1] means natural). Based on IA and IN scores, the photos are ranked from the highest score to the lowest score. According to IQ rank, the highest and the lowest quality photos of the scene are determined. The correlation between IQ and IA is evaluated by considering two questions: “is the IA rank of the highest IQ photo higher than that of the lowest one?” and “is the IQ rank of the highest IA photo higher than that of the lowest one?”. In a similar way, the correlation between IQ and IN is estimated by answering two sub questions: “is the IN rank of the highest IQ photo higher than that of the lowest one?” and “is the IQ rank of the highest IN photo higher than that of the lowest one?”. The correlations between IQ and IA and the ones between IQ and IN are evaluated for the 10 HDR scenes with and without reference and results are presented in Table 2 in which the correlation score is calculated as the number of correlated cases (the higher IQ rank, the higher IN rank or the higher IA rank). It appears that in the case of NRIQA (experiment without reference), the correlation between IQ and IA is much higher than that between

Table 2. Correlation between IQ and IA, IN in the two cases FRIQA and NRIQA. IQ vs. IN: “Is the IN rank of the highest IQ photo higher than that of the lowest one?”. IN vs. IQ: “Is the IQ rank of the highest IN photo higher than that of the lowest one?”. IQ vs. IA: “Is the IA rank of the highest IQ photo higher than that of the lowest one?”. IA vs. IQ: “Is the IQ rank of the highest IA photo higher than that of the lowest one?”. Y: Yes. N: No. score means correlation score.

			Scenes										
Correlation			1	2	3	4	5	6	7	8	9	10	Score
NRIQA	between	IQ vs. IN	Y	N	Y	Y	N	N	Y	N	N	N	8
	IQ and IN	IN vs. IQ	Y	N	Y	Y	N	N	Y	N	N	N	
	between	IQ vs. IA	Y	Y	Y	Y	N	Y	N	Y	Y	Y	15
	IQ and IA	IA vs. IQ	N	N	Y	Y	Y	Y	Y	Y	N	Y	
FRIQA	between	IQ vs. IN	Y	N	Y	N	N	N	Y	N	N	N	7
	IQ and IN	IN vs. IQ	Y	N	Y	N	N	Y	Y	N	N	N	
	between	IQ vs. IA	Y	Y	N	Y	Y	N	Y	N	Y	N	10
	IQ and IA	IA vs. IQ	Y	Y	N	N	Y	N	N	Y	N	N	

436 IQ and IN (correlation score: 15 versus 8). In contrast, in the case of FRIQA (experiment with reference),
 437 the difference between the two correlations decreases. The score of the correlation between IQ and IN
 438 is 7 while the score of the one between IQ and IA is 10. It seems that the influence of IA and IN on
 439 IQ is not the same for both NRIQA and FRIQA cases. The quality of experience might be the main
 440 cause of the differences. Specifically, IQ perception is affected by the visualization of the reference - a
 441 contextual factor of quality of experience. Without reference, viewers’ preference is mainly based on
 442 aesthetic perception. In contrast, IN has more influence on IQ than IA in the case of FRIQA.

443 Table 2 presents the correlation based on the comparison of IQ, IA and IN ranks between the
 444 highest IQ, IA, IN photos (the 1st photo in the lists ordered by IQ, IA, IN) and the lowest IQ, IA, IN
 445 photos (the 9th photo in the ordered lists) of nine versions generated from each HDR scene. Additional
 446 comparisons between the 2nd, 3rd highest IQ, IA, IN photos (the 2nd and 3rd photos in the ordered
 447 lists) and the 2nd, 3rd lowest IQ, IA, IN photos (the 7th and 8th photos in the ordered lists) are made
 448 to validate the assumption about the influence of IN and IA on IQ perception with and without
 449 reference. Correlations between IQ, IA and the ones between IQ, IN are estimated based on those
 450 comparisons and Pearson correlation between IQ, IA, IN ranks. Obtained results are presented in
 451 Table 3. Experiment results prove that the assumption is true: Without reference, IQ is more similar to
 452 IA than to IN while with reference, the roles of IN and IA in IQ perception are more balanced.

453 It could be explained by the fact that in the case of NRIQA, viewers’ decisions are generally made
 454 based on individual feelings, perception. In fact, viewers tend to pay more attention on high aesthetic
 455 photos than on low aesthetic photos. Besides, the lack of reference makes difficult for viewers to assess
 456 IN since they have to use feelings and memory to evaluate IN. Therefore, they might focus on IA - an
 457 easier aspect to assess IQ in the case of NRIQA (correlation score: 117 versus 82 and Pearson correlation:
 458 0.180 versus -0.070 for the correlation between IQ, IA and the one between IQ, IN respectively). In
 459 contrast, with reference, viewers share their attention on both IA criteria and the similarity between
 460 the compared versions and the reference. Viewers’ preference is not only affected by visual aesthetic
 461 perception referring to IA but also by technical factors (visible artifacts, obvious differences) referring
 462 to IN. It explains why the correlation between IQ, IN and the one between IQ and IA are more balanced
 463 in the case of FRIQA (correlation score: 94 vs 90 and Pearson correlation: 0.032 versus 0.028 for the
 464 correlation between IQ, IN and the one between IQ, IA respectively).

465 5. Conclusions

466 There are two main contributions in this paper related to the correlations between IA and IN in the
 467 relation with IQ. Firstly, the relations between IA and IN were investigated. Experiment results prove

Table 3. Correlations between IQ and IA, IN in the two cases FRIQA and NRIQA.

Comparison between	NRIQA		FRIQA	
	IQ and IN	IQ and IA	IQ and IN	IQ and IA
1st vs. 9th	8	15	7	10
1st vs. 8th	9	10	9	9
1st vs. 7th	9	14	10	10
2nd vs. 9th	10	17	9	13
2nd vs. 8th	7	13	13	13
2nd vs. 7th	8	16	11	10
3rd vs. 9th	10	11	9	10
3rd vs. 8th	11	11	15	7
3rd vs. 7th	10	10	11	8
Total	82	117	94	90

Pearson correlation between			
IQ and IN ranks	IQ and IA ranks	IQ and IN ranks	IQ and IA ranks
-0.070	0.180	0.032	0.028

468 that the correlation of IA features to IN features is quite significant but the correlation of IN features to
 469 IA features is much lower. Additionally, obtained results reflect that a high aesthetic photo does not
 470 mean a natural photo and a natural photo is not always considered as a high aesthetic photo. Besides,
 471 IA and IN features do not help improving significantly the performances of INA and IAA respectively.
 472 As a conclusion, although there are few moderate correlations and overlaps between IA and IN, they
 473 are two different notions reflecting different aspects of IQ. Secondly, the influences of IA and IN on IQ
 474 were evaluated and it appears that those influences are not the same depending on the experimental
 475 context (FRIQA or NRIQA). The main cause of the differences is quality of experience since contextual
 476 factors change observers' preference. Experiment results refer to the fact that viewers' IQ perception
 477 is more related to IA than to IN in NRIQA since NRIQA is mostly based on individual feelings and
 478 visual aesthetic perception. In contrast, the influence of IN and IA on IQ is more balanced in FRIQA
 479 because FRIQA is related to both individual opinions and technical aspects (technical errors, artifacts
 480 and specific screen) reflecting the similarity between transformed versions and original versions.

481 According to the current results, the direction of our research in the future is to develop algorithms
 482 able to enhance IQ based on both IA and IN aspects. Studying IA, IN and analyzing the correlation
 483 between IQ, IA and IN to understand positive influence and negative effects on IQ could be considered
 484 as the first step. The second step will be an improvement of IQA performance by considering IA and
 485 IN aspects. These two steps are the base to develop methods improving IQ by restoring naturalness of
 486 detected unnatural images and enhancing aesthetic quality of detected low aesthetic images.

487 Author Contributions:

488 **Quy-tien LE:** Conceptualization, Methodology, Visualization, Investigation, Software, Validation, Writing
 489 Original Draft, Editing.

490 **Patricia LADRET:** Supervision, Formal analysis, Conceptualization, Methodology, Reviewing.

491 **Huu-Tuan NGUYEN:** Supervision, Software, Reviewing.

492 **Alice CAPLIER:** Supervision, Formal analysis, Conceptualization, Methodology, Reviewing and Editing.

493 **Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal
 494 relationships that could have appeared to influence the work reported in this paper. The authors declare the
 495 following financial interests/personal relationships which may be considered as potential competing interests.

496 Abbreviations

497 The following abbreviations are used in this manuscript:

498

CNN	Convolutional Neural Network
FRIQA	Full Reference Image Quality Assessment
IA	Image Aesthetic
IAA	Image Aesthetic Assessment
499 IN	Image Naturalness
INA	Image Naturalness Assessment
IQ	Image Quality
IQA	Image Quality Assessment
NRIQA	No Reference Image Quality Assessment

500 References

- 501 1. LE, Q.T.; Ladret, P.; Nguyen, H.T.; Caplier, A. Study of naturalness in tone-mapped images. *Computer*
502 *Vision and Image Understanding* **2020**, *196*. doi:10.1016/j.cviu.2020.102971.
- 503 2. Zhai, G.; Zhang, W.; Yang, X.; Xu, Y. Image quality assessment metrics based on multi-scale edge
504 presentation. *IEEE Workshop on Signal Processing Systems Design and Implementation*, 2005. IEEE, 2005,
505 pp. 331–336.
- 506 3. Seshadrinathan, K.; Bovik, A.C. Unifying analysis of full reference image quality assessment. 2008 15th
507 IEEE International Conference on Image Processing. IEEE, 2008, pp. 1200–1203.
- 508 4. Wang, B.; Wang, Z.; Liao, Y.; Lin, X. HVS-based structural similarity for image quality assessment. 2008
509 9th International Conference on Signal Processing. IEEE, 2008, pp. 1194–1197.
- 510 5. Gao, X.; Lu, W.; Tao, D.; Li, X. Image quality assessment and human visual system. *Visual Communications*
511 *and Image Processing 2010*. International Society for Optics and Photonics, 2010, Vol. 7744, p. 77440Z.
- 512 6. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE*
513 *Journal of Oceanic Engineering* **2015**, *41*, 541–551.
- 514 7. Kim, J.; Lee, S. Deep learning of human visual sensitivity in image quality assessment framework.
515 Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1676–1684.
- 516 8. Khan, I.R.; Rahardja, S.; Khan, M.M.; Movania, M.M.; Abed, F. A tone-mapping technique based on
517 histogram using a sensitivity model of the human visual system. *IEEE Transactions on Industrial Electronics*
518 **2018**, *65*, 3469–3479.
- 519 9. Chen, Z.; Jiang, T.; Tian, Y. Quality assessment for comparing image enhancement algorithms. Proceedings
520 of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3003–3010.
- 521 10. Krasula, L.; Narwaria, M.; Fliegel, K.; Le Callet, P. Influence of HDR reference on observers preference
522 in tone-mapped images evaluation. 2015 Seventh International Workshop on Quality of Multimedia
523 Experience (QoMEX). IEEE, 2015, pp. 1–6.
- 524 11. Ke, Y.; Tang, X.; Jing, F. The design of high-level features for photo quality assessment. 2006 IEEE
525 Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, 2006, Vol. 1,
526 pp. 419–426.
- 527 12. Luo, Y.; Tang, X. Photo and video quality evaluation: Focusing on the subject. *European Conference on*
528 *Computer Vision*. Springer, 2008, pp. 386–399.
- 529 13. Leisti, T.; Radun, J.; Virtanen, T.; Halonen, R.; Nyman, G. Subjective experience of image quality: attributes,
530 definitions, and decision making of subjective image quality. *Image Quality and System Performance VI*.
531 International Society for Optics and Photonics, 2009, Vol. 7242, p. 72420D.
- 532 14. Luo, W.; Wang, X.; Tang, X. Content-based photo quality assessment. 2011 International Conference on
533 Computer Vision. IEEE, 2011, pp. 2206–2213.
- 534 15. Jayaraman, D.; Mittal, A.; Moorthy, A.K.; Bovik, A.C. Objective quality assessment of multiply distorted
535 images. 2012 Conference record of the forty sixth asilomar conference on signals, systems and computers
536 (ASILOMAR). IEEE, 2012, pp. 1693–1697.
- 537 16. Tang, X.; Luo, W.; Wang, X. Content-based photo quality assessment. *IEEE Transactions on Multimedia* **2013**,
538 *15*, 1930–1943.
- 539 17. Gu, K.; Wang, S.; Zhai, G.; Ma, S.; Yang, X.; Lin, W.; Zhang, W.; Gao, W. Blind quality assessment
540 of tone-mapped images via analysis of information, naturalness, and structure. *IEEE Transactions on*
541 *Multimedia* **2016**, *18*, 432–443.

- 542 18. Yaacoub, C.E.; Melhem, J.A.; Bilane, P.A. A no-reference metric for quality assessment of tone-mapped
543 high dynamic range images. *International Journal of Applied Engineering Research* **2017**, *12*, 2598–2603.
- 544 19. Jiang, G.; Song, H.; Yu, M.; Song, Y.; Peng, Z. Blind tone-mapped image quality assessment based on
545 brightest/darkest regions, naturalness and aesthetics. *IEEE Access* **2017**, *6*, 2231–2240.
- 546 20. Hii, Y.L.; See, J.; Kairanbay, M.; Wong, L.K. Multigap: Multi-pooled inception network with text
547 augmentation for aesthetic prediction of photographs. 2017 IEEE International Conference on Image
548 Processing (ICIP). IEEE, 2017, pp. 1722–1726.
- 549 21. Hosu, V.; Lin, H.; Sziranyi, T.; Saupe, D. KonIQ-10k: An ecologically valid database for deep learning of
550 blind image quality assessment. *IEEE Transactions on Image Processing* **2020**, *29*, 4041–4056.
- 551 22. Varga, D. Multi-pooled inception features for no-reference image quality assessment. *Applied Sciences* **2020**,
552 *10*, 2186.
- 553 23. Varga, D. Analysis of Benford’s Law for No-Reference Quality Assessment of Natural, Screen-Content,
554 and Synthetic Images. *Electronics* **2021**, *10*, 2378.
- 555 24. Li, Q.; Wang, Z. Reduced-reference image quality assessment using divisive normalization-based image
556 representation. *IEEE journal of selected topics in signal processing* **2009**, *3*, 202–211.
- 557 25. Rehman, A.; Wang, Z. Reduced-reference image quality assessment by structural similarity estimation.
558 *IEEE transactions on image processing* **2012**, *21*, 3378–3389.
- 559 26. Wu, J.; Lin, W.; Shi, G.; Li, L.; Fang, Y. Orientation selectivity based visual pattern for reduced-reference
560 image quality assessment. *Information Sciences* **2016**, *351*, 18–29.
- 561 27. Tang, L.; Sun, K.; Liu, L.; Wang, G.; Liu, Y. A reduced-reference quality assessment metric for
562 super-resolution reconstructed images with information gain and texture similarity. *Signal Processing: Image Communication* **2019**, *79*, 32–39.
- 564 28. Dhar, S.; Ordonez, V.; Berg, T.L. High level describable attributes for predicting aesthetics and
565 interestingness. CVPR 2011. IEEE, 2011, pp. 1657–1664.
- 566 29. Marchesotti, L.; Perronnin, F.; Larlus, D.; Csurka, G. Assessing the aesthetic quality of photographs using
567 generic image descriptors. 2011 international conference on computer vision. IEEE, 2011, pp. 1784–1791.
- 568 30. Mavridaki, E.; Mezaris, V. A comprehensive aesthetic quality assessment method for natural images using
569 basic rules of photography. 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015,
570 pp. 887–891.
- 571 31. Aydin, T.O.; Smolic, A.; Gross, M. Automated aesthetic analysis of photographic images. *IEEE Transactions*
572 *on Visualization and Computer Graphics* **2015**, *21*, 31–42.
- 573 32. Tian, X.; Dong, Z.; Yang, K.; Mei, T. Query-dependent aesthetic model with deep learning for photo quality
574 assessment. *IEEE Transactions on Multimedia* **2015**, *17*, 2035–2048.
- 575 33. Lu, X.; Lin, Z.; Jin, H.; Yang, J.; Wang, J.Z. Rating image aesthetics using deep learning. *IEEE Transactions*
576 *on Multimedia* **2015**, *17*, 2021–2034.
- 577 34. Wang, W.; Zhao, M.; Wang, L.; Huang, J.; Cai, C.; Xu, X. A multi-scene deep learning model for image
578 aesthetic evaluation. *Signal Processing: Image Communication* **2016**, *47*, 511–518.
- 579 35. LE, Q.T.; Ladret, P.; Nguyen, H.T.; Caplier, A. Image aesthetic assessment based on image classification
580 and region segmentation. *Journal of imaging* **2021**.
- 581 36. de Ridder, H.; Blommaert, F.J.; Fedorovskaya, E.A. Naturalness and image quality: chroma and hue
582 variation in color images of natural scenes. *Human Vision, Visual Processing, and Digital Display VI*.
583 International Society for Optics and Photonics, 1995, Vol. 2411, pp. 51–62.
- 584 37. de Ridder, H. Naturalness and image quality: saturation and lightness variation in color images of natural
585 scenes. *Journal of imaging science and technology* **1996**, *40*, 487–493.
- 586 38. Choi, S.Y.; Luo, M.; Pointer, M.; Rhodes, P. Investigation of large display color image appearance—III:
587 Modeling image naturalness. *Journal of Imaging Science and Technology* **2009**, *53*, 31104–1.
- 588 39. Cadik, M.; Slavik, P. The Naturalness of Reproduced High Dynamic Range Images. Proceedings of the
589 Ninth International Conference on Information Visualisation; IEEE Computer Society: Washington, DC,
590 USA, 2005; IV ’05, pp. 920–925.
- 591 40. Chollet, F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE
592 conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

- 593 41. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image
594 recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp.
595 8697–8710.
- 596 42. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam,
597 H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*
598 *arXiv:1704.04861* 2017.
- 599 43. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for
600 computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016,
601 pp. 2818–2826.
- 602 44. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv*
603 *preprint arXiv:1409.1556* 2014.
- 604 45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE*
605 *conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- 606 46. LE, Q.T.; Ladret, P.; Nguyen, H.T.; Caplier, A. Large Field/Close-Up Image Classification: From
607 Simple to Very Complex Features. *Computer Analysis of Images and Patterns*; Vento, M.;
608 Percannella, G., Eds. Springer, 2019, Vol. 11679, *Lecture Notes in Computer Science*, pp. 532–543.
609 *Proceedings of the 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Part II*,
610 doi:10.1007/978-3-030-29891-3_47.
- 611 47. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise reduction in speech*
612 *processing*; Springer, 2009; pp. 1–4.
- 613 48. Myers, L.; Sirois, M.J. Spearman correlation coefficients, differences between. *Encyclopedia of statistical*
614 *sciences* 2004, 12.
- 615 49. Union, I.T. *ITU-T Recommendation E.800. Definitions of terms related to quality of service*; 2008; p. 15.
- 616 50. Klima, M.; Fliegel, K.; Kekrt, D.; Dostal, P.; Podgorny, R. Image quality and QoE in multimedia systems.
617 2009 19th International Conference Radioelektronika. IEEE, 2009, pp. 3–10.
- 618 51. Brunnström, K.; Beker, S.A.; De Moor, K.; Doms, A.; Egger, S.; Garcia, M.N.; Hossfeld, T.; Jumisko-Pyykkö,
619 S.; Keimel, C.; Larabi, M.C.; Lawlor, B.; Le Callet, P.; Möller, S.; Pereira, F.; Pereira, M.; Perkis, A.; Pibernik,
620 J.; Pinheiro, A.; Raake, A.; Reichl, P.; Reiter, U.; Schatz, R.; Schelkens, P.; Skorin-Kapov, L.; Strohmeier, D.;
621 Timmerer, C.; Varela, M.; Wechsung, I.; You, J.; Zgank, A. Qualinet White Paper on Definitions of Quality
622 of Experience, 2013. Qualinet White Paper on Definitions of Quality of Experience Output from the fifth
623 Qualinet meeting, Novi Sad, March 12, 2013.
- 624 52. Drago, F.; Myszkowski, K.; Annen, T.; Chiba, N. Adaptive logarithmic mapping for displaying high
625 contrast scenes. *Computer graphics forum*. Wiley Online Library, 2003, Vol. 22, pp. 419–426.
- 626 53. Kuang, J.; Johnson, G.M.; Fairchild, M.D. iCAM06: A refined image appearance model for HDR image
627 rendering. *Journal of Visual Communication and Image Representation* 2007, 18, 406–414.
- 628 54. Mantiuk, R.; Myszkowski, K.; Seidel, H.P. A perceptual framework for contrast processing of high dynamic
629 range images. *ACM Transactions on Applied Perception (TAP)* 2006, 3, 286–308.
- 630 55. Mai, Z.; Mansour, H.; Mantiuk, R.; Nasiopoulos, P.; Ward, R.; Heidrich, W. Optimizing a tone curve for
631 backward-compatible high dynamic range image and video compression. *IEEE transactions on image*
632 *processing* 2010, 20, 1558–1571.