Which is the better inpainted image? Learning without subjective annotation

Mariko Isogawa¹ mariko.isogawa.kt@hco.ntt.co.jp Dan Mikami¹² dan.mikami.vp@hco.ntt.co.jp Kosuke Takahashi¹ kosuke.takahashi.rd@hco.ntt.co.jp Hideaki Kimata¹ hideaki.kimata.yu@hco.ntt.co.jp

- ¹ Nippon Telegraph and Telephone Corp. NTT Media Intelligence Laboratories
 1-1 Hikarino-oka, Yokosuka, Kanagawa, Japan
- ² Nippon Telegraph and Telephone Corp. NTT Communication Science Laboratories
 3-1 Morinosato-wakamiya, Atsugi, Kanagawa, Japan

Abstract

This paper proposes a learning-based quality evaluation framework for inpainted results that does not require any subjectively annotated training data. Image inpainting, which removes and restores unwanted regions in images, is widely acknowledged as a task whose results are quite difficult to evaluate objectively. Thus, existing learningbased image quality assessment (IQA) methods for inpainting require subjectively annotated data for training. However, subjective annotation requires huge cost and subjects' judgment occasionally differs from person to person in accordance with the judgment criteria. To overcome these difficulties, the proposed framework uses simulated failure results of inpainted images whose subjective qualities are controlled as the training data. This approach enables preference order between pairwise inpainted images to be successfully estimated even if the task is quite subjective. To demonstrate the effectiveness of our approach, we test our algorithm with various datasets and show it outperforms state-of-the-art IQA methods for inpainting.

1 Introduction

However, evaluating inpainted images is widely acknowledged as a task that can only be done by subjective judgment, and there is no definitive way of estimating subjective quality on the basis of objectively measurable features. One primary reason is illustrated in



Figure 1: An example that explains difficulty of evaluating inpainted images objectively.

Fig. 1. Both Figs. 1(c) and 1(d) are inpainted images with the same damaged region shown in Fig. 1(b). Although neither of these results are different from the original one shown in (a), both results are perceptually natural. In many studies it is considered that these results are "correct" as long as they are perceptually natural for humans, even if they differ from the original one [III, III, III]. However, this kind of "correctness" is defined only with subjective judgment. Because of this difficulty, estimating inpainting quality has long been considered a challenging task.

To tackle this problem, existing image quality assessment (IQA) methods have tried to find a way to represent subjective quality by means of objectively measurable indicators. Venkatesh and Cheung used observed gaze density inside and outside the damaged region in inpainted images [23]. Instead of observed gaze, many IQA methods using a computational visual saliency map, which simulates human gaze density, have also been proposed [3, 9, 13, 21]. However, actual human gazes vary by viewer and viewing context and their correspondence with saliency maps is quite limited. In addition, Isogawa et al. [15] revealed that the pixel-wise unnaturalness that occurs in inpainted images is not suitable for saliency based methods because the resolution of visual saliency maps is coarse.

Currently, learning based approaches are demonstrating their effectiveness [**D**, **LS**, **ZJ**]. Frantc et al. [**D**] and Voronin et al. [**ZJ**] proposed support vector regression (SVR) based approaches. Isogawa et al. [**LS**] achieved accurate estimation of subjective quality of inpainted images by dividing the problem into a set of pairwise preference order estimation tasks and used the learning-to-rank approach to solve them. These learning based methods commonly require a subjective annotation step before the training step, which is considered essential. This labor-intensive annotation leads to both huge annotation cost and, what is worse, fluctuation of evaluation criteria.

To overcome this problem, this paper proposes a new framework for learning-to-rank based preference order estimation with automatically generated training data. Throughout this paper we call such data "auto-generated" data. It simulates "failed" inpainting and assumes that a simulated sample has worse subjective quality than the method's best inpainted image. Thus, it generates training pairs automatically without any user intervention.

The main contribution of this paper is that the proposed method achieves learning based preference estimation of inpainted images without annotated training data. To the best of our knowledge, this is the first study that tackles learning based estimation of subjective attribute only with auto-generated data. The other contribution is the way of generating degraded inpainted results. The rest of this paper is organized as follows. In Section 2 we review related work. Section 3 describes the learning based ranking method we propose, which is trained with auto-generated data. Section 4 verifies the effectiveness of the proposed method. Section 5 concludes the paper with a summary of key points and mentions future work.

2 Related work

This section introduces related work. First, 2.1 overviews existing IQA methods for image inpainting. In 2.2, we focus on machine learning especially on the way of preparing training data. Studies on automatic generation or augmentation of training data are introduced.

2.1 Learning based IQA methods for inpainted image

Aiming at selecting the best one from a plurality of results among varied inpainted images, many IQA methods have been proposed [3, 9, 13, 13, 13, 14]. Among these methods, learning based approaches demonstrate effective performance [**B**, **I**, **D**]. Frantc et al. [**D**], and Voronin et al. [24] proposed SVR based IQA methods. These approaches estimate an absolute subjective score for each test image. For training the regression model, subjectively annotated rating scores are essential. Thus, they use data annotated by subjects who are asked to provide scores on a 5-point scale. Another learning based approach to tackle this problem is the learning-to-rank approach. It learns and estimates rank order on the basis of trained ranking function. The important advantage of this approach is that it can learn only on the basis of rank order. Because of this feature, this approach has been the focus of considerable attention, especially when it is applied to tasks that are difficult to estimate objectively [1], 6, 17, 17]. Isogawa et al. [15] proposed a learning-to-rank based IQA method by pairwise preference estimation. This method focuses on the premise that the preference order of inpainted images, rather than their absolute scores, is important for selecting the best one from a plurality of results, which is the primary goal of their method. For training data, the method requires image pairs with annotated preference order.

As described above, one difficulty that commonly exists in learning based methods is the need for a labor-intensive annotation step for obtaining training data. These manual annotations require huge annotation cost. In addition, the judgment criteria of subjects fluctuate occasionally. To overcome these problems, the proposed method enables automatic generation of training data. It generates pairwise training data automatically and applies a learning-to-rank based algorithm to the preference order estimation.

2.2 Learning with auto-generated training set

Incrementing the amount of training data generally leads to improved performance in machine learning. Therefore, in recent years some studies have improved learning accuracy by incrementing the learning data with automatic generation [19, 20]. Pishchulin et al. [19] proposed a human detection and pose estimation by using automatically generated training sets. The main advantage of their method is that it enables human poses and shapes to be controlled explicitly on the basis of existing training sets. They also combine various background images to increase training data. Ros et al. [20] proposed learning based pixel-wise semantic segmentation that uses automatically generated training data. Since annotation data for this task must be provided on a pixel-by-pixel basis, having humans provide the data is labor-intensive. To overcome the problem, they use realistic synthetic images of urban views in a virtual world that can provide annotation data on a pixel-by-pixel basis.

In these tasks, it is apparent that an image generation model can be obtained from annotation data. Thus, generation of training examples, i.e., a set of annotated data and the generated images, is rather easy. In contrast, modeling the relationship between inpainted images and their annotated subjective quality is quite difficult. The reason is shown in Fig. 1,



Figure 2: Overview of our proposed learning framework

where both inpainted images have subjectively good quality. This makes it difficult to create the auto-generated training data. To the best of our knowledge, this paper is the first trial of making auto-generated training data for subjectively assessing inpainted image quality.

3 Proposed method

3.1 Overview

Figure 2 shows the overview of the proposed preference ordering framework with autogenerated training set. Our framework consists of training and estimation phases.

In the training phase, the proposed method first generates a simulated training set. Then, a ranking model is trained with these auto-generated images. As the ranking model, we utilize the pairwise learning-to-rank based method previously used by Isogawa et al. [13]. This is because unlike SVR-based methods [1, 23], a pairwise method only requires a set of preference orders between two images and does not require absolute scores of subjective quality. The learning process is detailed in 3.2. Because the proposed method uses pairwise learning-to-rank, the proposed generation of training data yields inpainted image pairs with known preference orders. This is described in more detail in 3.3.

The estimation phase procedure is fairly clear; it compares the ranking function outputs from two images. The subsequent subsections mainly focus on the learning algorithm, which includes an algorithm for ranking and generating simulated images for training.

3.2 Model learning with auto-generated training data

This subsection describes a learning-to-rank based learning algorithm for ordering pairwise images. The method is trained with auto-generated simulation images that represent degraded inpainted results. This algorithm premises a ranking function f(x) that projects sample images to a one-dimensional axis in accordance with the subjective quality of inpainted results. Hereafter, we use x^i to denote a feature vector extracted from sample image I^i . For simplicity, we use " $I^i > I^j$ " to express that " I^i is preferred to I^j ".

For easy understanding, before we describe the ranking algorithm trained with autogenerated training data, let us briefly explain the algorithm trained with subjectively anno-



Figure 3: Typical patch-based approach for image inpainting

tated data. We define the function $h(x^i, x^j)$ that denotes human annotation as follows:

$$h(x^{i}, x^{j}) = \begin{cases} +1 \ (I^{i} \succ I^{j}) \\ 0 \ (\text{no preferences}) \\ -1 \ (I^{j} \succ I^{i}), \end{cases}$$
(1)

The f(x) is trained so that the difference of outputs $f(x^i) - f(x^j)$ has the same sign as $h(x^i, x^j)$. In a word, the function f should satisfy the following formula:

$$\operatorname{sign}(h(x^{i}, x^{j})) = \operatorname{sign}(f(x^{i}) - f(x^{j})).$$
(2)

The goal is to learn f which is concordant with the training samples. We modeled f with the linear function $f(x) = \omega^{\top} x$. Then Eq. 2 can be rewritten as

$$\operatorname{sign}(h(x^{i}, x^{j})) = \operatorname{sign}(\boldsymbol{\omega}^{\top}(x^{i} - x^{j})).$$
(3)

That is, to find the weight vector ω satisfying Eq. 3 for most training data pairs. This is the same problem as that of binary classification. We use a pairwise learning-to-rank algorithm called RankingSVM [\square] to solve it.

Now we are ready to introduce auto-generated training set into the ranking algorithm. Let I_{sim}^l be a simulated inpainted image with *l* degraded level, with which larger *l* indicates more degradation. In accordance with degraded level, the preference order among such images is

$$I^l_{sim} \succ I^m_{sim} \iff l < m. \tag{4}$$

Such auto-generated images are used to train the ranking function. The way to generate degraded images I_{sim} is described in the next subsection.

3.3 Automatic training data generation

The proposed method relies on existing inpainting methods and devices for them to obtain degraded inpainted images that well simulate inpainting failures. Subsection 3.3.1 briefly reviews typical inpainting algorithms and then 3.3.2 describes how degraded data are generated on the basis of the existing inpainting algorithms.

3.3.1 Patch based image inpainting algorithms

Among various inpainting methods, patch-based algorithms are widely acknowledged as promising approaches. Typically they comprise three steps, which we will explain by using Fig. 3. For the damaged region masked with red in Fig. 3(a), (1) a patch that includes both



Figure 4: Patch based degraded inpainted image generation.

the source and the damaged region is set as a target, the blue rectangle in Fig. 3(b), (2) a similar patch for the target patch, the green rectangle in Fig. 3(b), is retrieved in the source region, and (3) the damaged region in the target patch is replaced in accordance with similar patches as in Fig. 3(c). The resultant restored image is shown in Fig. 3(d).

Here, P(p) is a target patch whose center pixel is p, and \hat{P}_p^N denotes the *N*-th most similar patch from P(p). Since total inpainting quality highly depends on the quality of retrieved patches, it is widely acknowledged that retrieving the most similar patches provides the best inpainted quality. That is, for "fine" inpainted results, the most similar patch (i.e., with N = 1) from P(p), which is denoted as \hat{P}_p^1 , is retrieved as in Eq. 5 to restore a missing region.

$$\hat{P}_{p}^{1} = P(q') = \arg\min_{P(q)} dist(P(p), P(q))$$
(5)

 $dist(\cdot)$ represents distance function. The proposed method uses the assumption in an inverse way, i.e., dissimilar patches generate unnatural inpainted images.

3.3.2 Auto-generated inpainted image as a training set

In simulating failed inpainted images, we found that if we selected the *N*-th most similar patch \hat{P}^N having larger *N*, it would apparently correspond to the cases in which good patches for inpainting cannot be found. This is a typical case of inpainting failure. Therefore, as the value *N* gets larger, the patch becomes dissimilar and the completion results get worse. That is, simulated inpainted images are generated so that their relationships depend on the level of similar patches as $I_{sim}^{l1} \succ I_{sim}^{l2}$ when l1 < l2, where I_{sim}^N represents a simulated image inpainted with *N*-th similar patches. We propose incorporating this patch retrieval into existing inpainting algorithms. Here we show our simulated image generation with two base algorithms [**D**, **II**] as examples.

With Criminisi et al's method [2]. The original method uses simple patch retrieval as shown in Eq. 5. Our method can be easily incorporated into this patch retrieval; instead of retrieving the most similar patches \hat{P}_p^1 , we obtain \hat{P}_p^L with L > 1. Figure 4 demonstrates it in more detail. The damaged region and the corresponding inpainting target image are shown in (a) and (b). The resultant inpainted images are shown in (c) and (d). Here, (c) is the method's best result and (d) is a simulated deteriorated result obtained using our proposed method. As the typical procedure of patch-based inpainting, target patch P(p) is determined

in (i), and then a similar patch is retrieved and used for filling in the hole. In case (c), the most similar patch is used. Unlike this, the proposed method uses a dissimilar patch depending on degraded level L and obtains the degraded result (d-ii).

With He and Sun's method [11]. Some current studies extend the basic algorithms by using patch retrieval indirectly. He and Sun improve the basic algorithm on the basis of two ideas; extension of the patch *P* to the whole image *I*, and treating an inpainting task as a Photomontage [**D**] problem. Let \hat{I}^N be the *N*-th most similar image for damaged image *I*. They retrieve the *K* most similar images $\hat{I}_K = \{\hat{I}^1, \hat{I}^2, ..., \hat{I}^K\}$ and the missing region is filled by combining a stack of these images. Our method modifies this image retrieval part; instead of retrieving the *K* most similar images, we obtain $\hat{I}_{\alpha L}$ that excludes $\hat{I}_{\alpha(L-1)}$. That is, we obtain $\hat{I}_{\alpha L} - \hat{I}_{\alpha(L-1)} = \{\hat{I}^{\alpha(L-1)+1}, \hat{I}^{\alpha(L-1)+2}, ..., \hat{I}^{\alpha L}\}$ to generate an *L*-th level of a degraded image.

4 Experiment

In this section we report how we investigated the efficacy of the proposed method. Subsection 4.1 shows the experimental setup we used, including training data preparation. Subsection 4.2 compares the proposed method to existing IQA methods. For further analysis, we changed the setting of the automatic training data generation and analyzed how it affects the estimation performance in 4.3.

4.1 Experimental setup

4.1.1 Ranking learning and manual annotation

The proposed ranking system is trained with RankingSVM with SVM Rank [22] with Radial Basis Function (RBF) as the kernel function with tuned parameters. As image features that represent all data elements, we used those proposed by Isogawa et al. [13], which focus on the unnaturalness produced by color or structural discontinuity.

For subjectively annotated data used as training data and test data, we prepared the data in the following way. We used He and Sun's inpainting method [III] to inpaint 100 images with manually masked damaged regions. For each masked image, we generated six inpainted results with different parameters. We controlled two parameters; one is the patch size, having three options; the other is the number of similar images to be retrieved, having two options. In total, we got six options by multiplying them. The quality of the images was evaluated by 8 subjects (4 males and 4 females) with normal vision. To make the users' judgment easy, we randomly displayed a pair of inpainted images side-by-side. Subjects were asked to choose one of three options: right image is better, left image is better, and no preference order (i.e., it is hard to decide which one is better or which one is worse). These manually annotated data were divided into training data and test data.

We denote T_a^S to specify these annotated data sets for which a consensus was obtained for at least *S* subjects, e.g., T_a^8 indicates the pairs that got a unanimous answer from all 8 subjects, and T_a^5 contains that having consensus from 5, 6, 7 or 8 subjects. Thus, more than half of the subjects gave the same preference order to pairs in T_a^5 . Therefore, *S* reflects the difficulty humans have in making judgments in such cases. Excluding inpainted images with extremely poor quality, the amounts of test data $|T_a^S|$ of T_a^S with S = 5 to 8 were $(|T_a^5|, |T_a^6|, |T_a^7|, |T_a^8|) =$ (184, 136, 71, 38). Following 4.2 and 4.3, we investigated how methods effectively estimate these subjectively annotated pairwise preference orders.



with damaged region (b) Close-up view of automatically generated training images with degraded level L

Figure 5: Simulated inpainted images.

4.1.2 Auto-generated training data

Because He and Sun's method was used for inpainting, we applied the proposed training sample generation method to He and Sun's inpainting algorithm [11], and augmented the training data. We used the same inpainting algorithm with test data, with He and Sun's method [11] as a base algorithm for a simulation. 100 images were masked with a pseudo damaged region. Note that since the position, size, and shape of the damaged region are normalized during the learning-to-rank process, we are able to set the damaged region arbitrary regardless of objects in the target images.

We set five degradation levels for simulating inpainted images I_{sim}^L , i.e., L = 0, 1, 2, 3, 4. L = 0 indicates an image with no degradation; it is equal to the method's best inpainted image. These five image levels are generated from one original image. By obtaining combinations of these five images, ${}_{5}C_{2} = 10$ image pairs are provided. These image pairs have pseudo preference orders $I_{sim}^x \succ I_{sim}^y$ ($\forall x < y$). We denote such auto-generated degraded images for training T_d .

Figure 5(b) shows degraded images depending on the degraded level L (L = 0, 1, 2, 3, 4). All degraded images are inpainted with the masked region shown in (a). Figure 5 shows that our method simulates degraded images well; each degraded image gets worse quality as Lincreases. Though it is subjective, this deterioration well simulates the failures that typically occur in ordinary inpainting methods having inappropriate parameters such as patch size.

4.2 Comparison with existing methods

We conducted experiments comparing our method to other IQA methods for image inpainting, i.e., ASVS and DN by Ardis and Singhal. [3], \overline{GD}_{in} by Venkatesh and Cheung [23], BorSal, StructBorSal by Oncu et al. [13] as non learning-based methods, and Isogawa et al.'s method [13] as a learning-based method. Note that although the \overline{GD}_{in} originally uses measured human gaze, we used a saliency map instead. This is the same evaluation approach used in Oncu et al. [13]. For training Isogawa et al.'s method we used the annotation data of T_a^8 . Our proposed learning method trained with auto-generated data T_d is denoted as $Ours(T_d)$.

Table 1 shows the prediction accuracy for all test data T_a^S ($5 \le S \le 8$) obtained for each metric. Our method *Ours*(T_d) correctly estimated the image pair preferences with the highest

Table 1: Prediction accuracy comparison with existing image quality assessment metrics [%]



score for all of test data; the improvement our method achieved over Isogawa et al.'s method was 6.52, 8.09, 4.23, 2.63 points for test data T_a^5 , T_a^6 , T_a^7 , and T_a^8 .

4.3 Effectiveness verification for auto-generated training data depending on varied conditions

This subsection describes three more investigations we conducted to verify the validity of the auto-generated data. The first one investigates the influence of volume of auto-generated data on estimation accuracy, which is the ratio of estimation success of preference order among annotated pairs. Fig. 6 shows estimation accuracy when the amount of auto-generated data is increased from 90 to 990 in 60 increments. Training data are randomly selected from T_d . As shown in the graphs in the figure, estimation accuracy increased as the amount of training data increased. In addition, to investigate whether auto-generated data can be used as a substitute for annotated data, the performances depending on the proportion of autogenerated data were tested. Fig. 7 shows performances depending on the proportion of autogenerated data with T_a^5 . Although T_a^5 is considered to be less reliable because of its lower consensus as we will discuss later, it is useful for verifying changes in performances relative to those in composition ratios of annotated and auto-generated training data because it has the largest number of annotated data. Here, the number of training data was fixed to 990 in all cases; only the proportion of annotated and auto-generated data was changed. The number of annotated data was decreased from 180 to 0 in 10 decrements. Even though the number of subjectively annotated training data was changed, the estimation accuracies were almost constant in all cases.

The second one investigates the influence of levels of auto-generated data on estimation accuracy. Fig. 8 shows estimation accuracy for each T_a^S when the auto-generated training data levels L are changed between L = 1 to 4. The size of available training data is respectively 99, 297, 594, and 990 for L = 1, 2, 3, and 4. However, to focus on the affect of data

Table 2: Prediction accuracy with orwithout subjectively annotated data [%]

Table 3: Prediction accuracy with orwithout unreliable data [%]

	T_a^5	T_a^6	T_a^7	T_a^8		T_a^5	T_a^6	T_a^7	T_a^8
$Ours(T_d)$	66.85	<u>70.59</u>	76.06	<u>78.95</u>	$Ours(T_d^{re})$	<u>66.30</u>	<u>69.85</u>	76.06	<u>81.58</u>
$Ours(T_d + T_a)$	65.22	68.38	76.06	<u>78.95</u>	$Ours(T_d^{un})$	63.04	67.65	73.24	78.95

levels, we set it to the smallest number 99, i.e., that for L = 1. Note that the results are average performances with 10 trial runs and that training data for each trial are randomly selected. As the figure shows, the estimation accuracy increases as *L* increases, which suggests that multi-levels of auto-generated data work effectively. These two kinds of investigations suggest that the auto-generated data works as expected.

For further investigation, the third one verifies the performance when subjectively annotated data is added to auto-generated data. We added T_a^S to T_d for further verification of auto-generated data performance. We denote our learning method with such data as $Ours(T_d + T_a)$. The comparison between $Ours(T_d)$ and $Ours(T_d + T_a)$ with all test data T_a^S with S = 5, 6, 7, 8 is shown in Table 2. With this table we found interesting results; $Ours(T_d)$ gets better results than $Ours(T_d + T_a)$ with T_a^5 and T_a^6 even though $Ours(T_d + T_a)$ has more training data. We consider that one possible reason is the reliability of subjectively annotated data. T_a^S with lower S is considered to be less reliable; the consensus about its reliability is low and this indicates it is difficult to judge even for humans. This assumption and prediction the accuracy results in Table 2 lead us to consider that unreliable data deteriorate total estimation accuracy.

To verify this consideration, we conducted further experiment on data reliability using auto-generated data. We divide the auto-generated data, T_d into two groups; reliable data set consisted by auto-generated data of L = 1 and 4, T_d^{re} ; unreliable data set consisted by that of L = 2 and 3, T_d^{un} . We subjectively confirmed that T_d^{un} have small difference in subjective quality and are difficult to be judged by a large margin compare to T_d^{re} , as in Fig. 5. The prediction accuracy is shown in Table 3. For all test data, $Ours(T_d^{re})$ excels $Ours(T_d^{un})$. It also suggests that subjectively similar data like T_d^{un} is not a good data for training. Thus, we should consider the balance between the number of auto-generated data and the quality of them as a future work.

5 Conclusion

This paper describes a learning-based ranking framework for image inpainting. Unlike existing learning-based IQA methods, our method trains the system without subjectively annotated data by using auto-generated data; we used simulated "failed" inpainted images by focusing on inpainting algorithms. Preference order estimation experiment results suggest the method's efficacy and several investigations suggest the validity of using auto-generated data instead of subjectively annotated data. In future work we will optimize the balance between the number of auto-generated data and the quality of them for our proposed system with a larger dataset. Also, we believe that the idea of generating training data by daringly generating failed images can be widely applied to other tasks requiring subjective evaluations such as image colorization [1] or image transfer [1]. Investigating the efficacy for these other tasks is also a subject for our future work.

References

- Takashi Abe, Takayuki Okatani, and Koichiro Deguchi. Recognizing surface qualities from natural images based on learning to rank. In *International Conference on Pattern Recognition (ICPR)*, pages 3712–3715, 2012.
- [2] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen. Interactive digital photomontage. *ACM Transactions on Graphics (TOG)*, 23(3):294–302, 2004.
- [3] Paul A. Ardis and Amit Singhal. Visual salience metrics for image inpainting. In *Proc. SPIE*, volume 7257, pages 72571W–72571W–9, 2009.
- [4] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics (TOG), 28(3), 2009.
- [5] Marcelo Bertalmio, Luminita Vese, Guillermo Sapiro, and Stanley Osher. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889, 2003.
- [6] Kuang-Yu Chang and Chu-Song Chen. A learning framework for age rank estimation based on face images with scattering transform. *IEEE Transactions on Image Processing*, 24(3):785–798, 2015.
- [7] Antonio Criminisi, Patrick Perez, and Kentaro Toyama. Region filling and object removal by exemplar-based inpainting. *IEEE Transactions on Image Processing*, 13(9): 1200–1212, 2004.
- [8] Soheil Darabi, Eli Shechtman, Connelly Barnes, Dan B Goldman, and Pradeep Sen. Image Melding: Combining inconsistent images using patch-based synthesis. ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH 2012), 31(4):82:1– 82:10, 2012.
- [9] V. A. Frantc, V. V. Voronin, V. I. Marchuk, A. I. Sherstobitov, S. Agaian, and K. Egiazarian. Machine learning approach for objective inpainting quality assessment. In *Proc. SPIE*, volume 9120, pages 91200S–91200S–9, 2014.
- [10] Kaiming He and Jian Sun. Image completion approaches using the statistics of similar patches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(12): 2423–2435, 2014.
- [11] Ralf Herbrich, Thore Graepel, and Klaus Obermayer. Large Margin Rank Boundaries for Ordinal Regression, chapter 7, pages 115–132. MIT Press, January 2000.
- [12] Aaron Hertzmann, Charles E. Jacobs, Nuria Oliver, Brian Curless, and David H. Salesin. Image analogies. In *Proc. ACM SIGGRAPH*, pages 327–340, 2001.
- [13] Jia-Bin Huang, Sing Bing Kang, Narendra Ahuja, and Johannes Kopf. Image completion using planar structure guidance. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014), 33(4):129:1–129:10, 2014.

- [14] Mariko Isogawa, Dan Mikami, Kosuke Takahashi, and Akira Kojima. Image and video completion via feature reduction and compensation. *Multimedia Tools and Applications*, pages 1–20, 2016. ISSN 1573-7721.
- [15] Mariko Isogawa, Dan Mikami, Kosuke Takahashi, and Akira Kojima. Eye gaze analysis and learning-to-rank to obtain the most preferred result in image inpainting. In *IEEE International Conference on Image Processing (ICIP)*, pages 3538–3542, 2016.
- [16] Aditya Khosla, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Memorability of image regions. In Advances in Neural Information Processing Systems (NIPS), pages 296–304, 2012.
- [17] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. In *Proc. ACM SIGGRAPH*, pages 689–694, 2004.
- [18] AlexandraIoana Oncu, Ferdinand Deger, and JonYngve Hardeberg. Evaluation of digital inpainting quality in the context of artwork restoration. In *European Conference on Computer Vision (ECCV) Workshops and Demonstrations*, volume 7583, pages 561– 570, 2012.
- [19] Leonid Pishchulin, Arjun Jain, Mykhaylo Andriluka, Thorsten Thormählen, and Bernt Schiele. Articulated people detection and pose estimation: Reshaping the future. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3178– 3185, 2012.
- [20] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3234–3243, 2016.
- [21] A.DANG Thanh Trung, B.Azeddine Beghdadi, and C.Chaker Larabi. Perceptual quality assessment for color image inpainting. In *IEEE International Conference on Image Processing (ICIP)*, pages 398–402, 2013.
- [22] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. Large margin methods for structured and interdependent output variables. *The Journal of Machine Learning Research*, 6:1453–1484, 2005. ISSN 1532-4435.
- [23] M. Vijay Venkatesh and Sen-Ching S. Cheung. Eye tracking based perceptual image inpainting quality analysis. In Proc. IEEE International Conference on Image Processing (ICIP), pages 1109–1112, 2010.
- [24] V. V. Voronin, V. A. Frantc, V. I. Marchuk, A. I. Sherstobitov, and K. Egiazarian. Noreference visual quality assessment for image inpainting. In *Proc. SPIE*, volume 9399, pages 93990U–93990U–8, 2015.
- [25] Zongben Xu and Jian Sun. Image inpainting by patch propagation using patch sparsity. *IEEE Transactions on Image Processing*, 19(5):1153–1165, 2010. ISSN 1057-7149.
- [26] Jianzhou Yan, Stephen Lin, Sing Bing Kang, and Xiaoou Tang. A learning-to-rank approach for image color enhancement. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2987–2994, 2014.