Towards Top-Down Just Noticeable Difference Estimation of Natural Images

Qiuping Jiang, Zhentao Liu, Shiqi Wang, Feng Shao, and Weisi Lin, *Fellow, IEEE*

***Abstract*—Just noticeable difference (JND) of natural images refers to the maximum pixel intensity change magnitude that typical human visual system (HVS) cannot perceive. Existing efforts on JND estimation mainly dedicate to modeling the diverse masking effects in either/both spatial or/and frequency domains, and then fusing them into an overall JND estimate. In this work, we turn to a dramatically different way to address this problem with a top-down design philosophy. Instead of explicitly formulating and fusing different masking effects in a bottom-up way, the proposed JND estimation model dedicates to first predicting a critical perceptual lossless (CPL) counterpart of the original image and then calculating the difference map between the original image and the predicted CPL image as the JND map. We conduct subjective experiments to determine the critical points of 500 images and find that the distribution of cumulative normalized KLT coefficient energy values over all 500 images at these critical points can be well characterized by a Weibull distribution. Given a testing image, its corresponding critical point is determined by a simple weighted average scheme where the weights are determined by a fitted Weibull distribution function. The performance of the proposed JND model is evaluated explicitly with direct JND prediction and implicitly with three applications including JND-guided noise injection, JND-guided image compression, and distortion detection and discrimination. Experimental results have demonstrated that promising performance of the proposed JND model. The data and code of this work are available at https://github.com/Zhentao-Liu/KLT-JND.**

***Index Terms*—Just noticeable difference, distortion visibility, masking effect, critical perceptual lossless, Karhunen-Loéve Transform.**

I. Introduction

Just noticeable difference (JND) refers to the maximum change magnitude that typical human visual system (HVS) cannot perceive [1]. It reveals the visibility limitation of the HVS and reflects the underlying perceptual redundancy in visual signals, rendering it useful in many perceptual image/video processing applications such as image/video compression [2–5], perceptual image/video enhancement [6, 7], information hiding and watermarking [8–14], and visual quality assessment [15–19], etc. Due to its wide applications, JND estimation of natural images has received much attention and been widely investigated. Generally, the JND models can be divided into two categories: JND models in pixel domain and JND models in frequency domain.

JND models in pixel domain directly calculate the JND at pixel level with considerations of either/both luminance adaption (LA) or/and contrast masking (CM) effects of the HVS. As the pioneering work, Chou *et al.* [20] first proposed a spatial-domain JND model by combining LA and CM. Afterwards, Yang *et al.* [21] further proposed a generalized spatial JND model with a nonlinear additivity model for masking effects (NAMM) to characterize the possible overlaps between LA and CM. Based on Yang *et al.*'s work, Liu *et al.* [22] introduced an enhanced pixel-level JND model with an improved scheme for CM estimation. Specifically, the image is first decomposed into structural component (*i.e.*, cartoon like, piecewise smooth regions with sharp edges) and textural component using a total-variation algorithm. Then, the structural and textural components are used estimating EM and TM effects, respectively. In order to differently manipulate order and disorder regions, Wu *et al.* [23] designed a novel JND estimation model based on the free-energy principle. An autoregressive model is first applied to predict the order and disorder contents of an input image, then different schemes are used to estimate the JND threshold of these two parts, respectively. Further, Wu *et al.* [24] took the concept of pattern complexity (PC) into account for JND estimation. They quantified the pattern complexity as the diversity of pixel orientations. Finally, pattern masking is deduced and combined with the traditional CM for JND estimation. Jakhetiya *et al.* [25] further combines RMS contrast with LA and CM to build a more comprehensive JND model in low-frequency regions. For high-frequency regions, a feedback mechanism is used to efficiently mitigate the over- and under-estimations of CM. Chen *et al.* [26] took horizontal-vertical anisotropy and vertical-meridian asymmetry into consideration to yield a better JND estimation. The basic consideration is that the effect of eccentricity on visual sensitivity is not homogeneous across the visual field. Shen *et al.* [27] decompose the image into three components namely luminance, contrast, and structure. Since the masking of structure visibility (SM) is unknown, they trained a deep learning-based structural degradation estimation model to approximate SM. Finally, LA, CM and SM are combined to estimate the overall JND. Wang *et al.* [28] proposed a novel JND estimation model by exploiting the hierarchical predictive coding theory. They

This work was supported in part by the Zhejiang Natural Science Foundation under Grant LR22F020002, in part by the Natural Science Foundation of China under Grants 61901236 and 62071261, and in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant SJLZ2020003. *(The first two authors contribute equally to this work. Corresponding author: Qiuping Jiang)*

Q. Jiang, Z. Liu, and F. Shao are with the School of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: jiangqiuping@nbu.edu.cn, zhentaoliu0319@163.com, shaofeng@nbu.edu.cn).

S. Wang is with the Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong (e-mail: shiqwang@cityu.edu.hk).

W. Lin is with the School of Computer Science and Engineering, Nanyang Technology University, Singapore (e-mail: wslin@ntu.edu.sg).

simulated both the positive and negative perception effects of each stage individually and integrated them with Yang's NAMM model [21] to yield the total JND threshold.

JND models in frequency domain firstly transform the original image into a specific transform domain and then the corresponding JND thresholds for each sub-band are estimated. The main consideration of these frequency domain-based JND models is to make use of the well-known contrast sensitivity function (CSF) which reflects the bandpass characteristics of HVS in the spatial frequency domain and is typically modeled as an exponential function of the spatial contrast [29]. In [30], the visibility thresholds for different frequencies are measured through subjective tests, and the CSF is built to account for the fundamental/base JND threshold for each sub-band. Typically, the JND thresholds for each sub-band are usually estimated based on a fixed size block (e.g., 8×8) via a linear combination of CSF and some other modulation factors. For example, Wei and Ngan [31] utilized a simple piecewise function to represent LA and formulated the CM as a categorizing function according to the richness of block texture information. Bae *et al.* [32] proposed a new DCT-based JND profile by incorporating the CSF, LA, and CM effects. Specifically, a new CM JND is modeled as a function of DCT frequency and a newly proposed structural contrast index (a new texture complexity metric that considers not only contrast intensity, but also structureness of image patterns). Wan *et al.* [33] analyzed orientation information with the DCT coefficients and a more accurate CM model was proposed in the DCT domain. Besides LA and CM, some works also considered foveated masking as influential factors for JND estimation. The foveated masking was first modeled by Chen *et al.* in [26] where it was utilized as an explicit form in pixel-domain JND and then incorporated in the frequency-domain JND estimation.

The above only provides a brief overview of exsiting advances in the field of JND estimation and a more comprehensive survey can be found in [1, 29]. In general, the existing JND models share the same design philosophy, i.e., modeling the visibility masking effects of different factors and then fusing them together to derive an overall JND estimation. Such a kind of design philosophy can be considered as a bottom-up strategy which starts from all possible concrete influential factors and then progressively produce the final estimation. However, this design philosophy has some inherent drawbacks. First, without having a deep understanding of the HVS properties, it is hard to take all influential factors into account. Second, the interactions among different masking effects are difficult to be characterized. Therefore, the performance of the existing JND estimation models remains limited.

In this paper, we turn to a different way to address the above problems according to a top-down perspective. Keeping in mind that the goal is to estimate the maximum change magnitude (JND threshold) of each pixel that typical HVS cannot perceive. In other words, with an ideal JND map, if the original image is changed within the JND threshold, we can still perceive the changed image as a perceptual lossless one. Thus, we intuitively come up with the idea to firstly determine a critical perceptual lossless (CPL) image and then calculate the difference between the CPL image and the original image to serve as the JND map, thus bypassing the inaccurate formulation and integration of different masking effects. The CPL image refers to a changed image whose change magnitude is just at the critical point beyond which this changed image will be either perceived as lossy or not the critical one (although perceptual lossless). To facilitate comparing the bottom-up and top-down design philosophies for JND estimation, we illustrate their pipelines in Fig. 1. As shown, our proposed top-down JND estimation (right column) first starts from the generation of a CPL image from the original one and then deduces the final JND map by calculating their difference. However, the traditional bottom-up JND estimation (left column) first starts from the formulation and integration of different masking effects to derive the overall JND map based on which a perceptual lossless image is expected to be obtained (the final result may not be perceptual lossless since the JND map established in this way may not be sufficiently accurate).

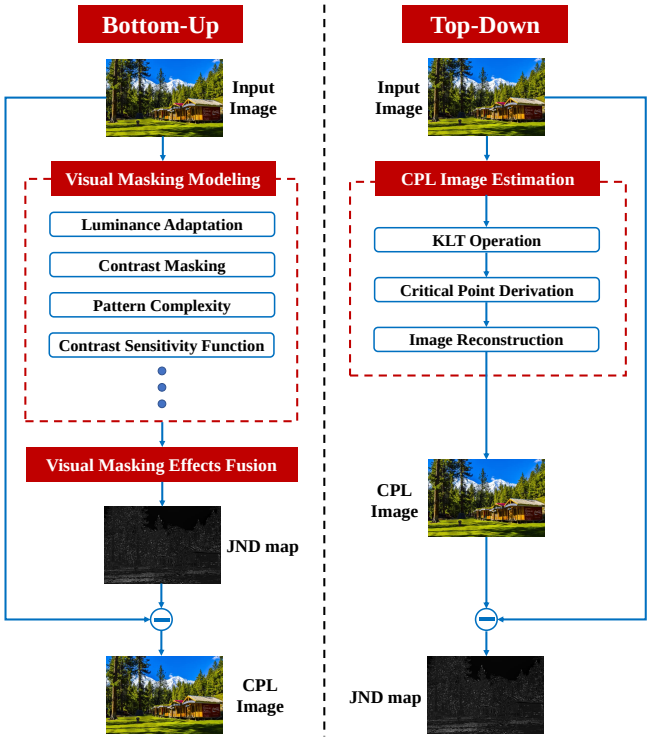


Fig. 1: Pipelines of bottom-up and top-down design philosophies for JND estimation.

Given a specific image as input, we first perform the Karhunen-Loéve Transform (KLT) and then derive its critical point (i.e., perceptual lossless threshold) by exploiting the convergence characteristics of KLT coefficient energy. Once the critical point is determined, the CPL image can be reconstructed. Then, the difference map between the original image and the CPL image is deemed as the visually redundant information that cannot be perceived by HVS, implying the visibility limitation of the HVS. Finally, we just simply take the derived difference map as the final JND map.

Although we have mentioned that one of the main shortcom-

ings of the current bottom-up JND estimation approaches is the insufficient understanding of the HVS properties, we do not intend to interpret all perceptual processes with KLT. Instead, we just try to address the JND estimation problem from another perspective by exploiting certain properties of KLT to identify the critical point (indicating how many spectral components are involved) for reconstructing a CPL image. Although the objective is to estimate JND map, the starting points of the conventional JND models and our proposed JND model are different. That is, the conventional JND models usually follow a bottom-up design philosophy which starts from modeling the visibility masking effects of different factors and then fusing them together to obtain a final JND estimation, while the proposed JND model is designed from another perspective, i.e., top-down perspective, which dedicated to determining the CPL as accurate as possible. Obviously, the modeling of masking effect of different influential factors requires substantial knowledge about the HVS properties while the determination of CPL is much simpler in this sense. In addition, we notice that there is some research on investigating the visually lossless threshold (VLT) of image compression [34, 35], which is similar to the critical point defined in our paper. However, compared with the widely used Quality Factor (QF) which is mainly applicable to image compression [34–36], the spectral component in KLT is not restricted to any specific applications. In the literature, there is also another concept, i.e., Visual Difference Predicator (VDP) [37], which is a relevant concept with JND. However, their definitions are different. The JND model and VDP metric have different input signals and their outputs also have different physical meanings. Let us denote the original image as I_o, the distorted image as I_d, the JND model as F_{JND}, the output of the JND model as O_{JND}, the VDP metric as F_{VDP}, and the output of the VDP metric as O_{VDP}. The mathematical definitions of a JND model and a VDP metric can be briefly expressed as follows:

$$O_{JND} = F_{JND}(I_o) \tag{1}$$

$$O_{VDP} = F_{VDP}(I_o, I_d) \tag{2}$$

According to the above mathematical formulations, we observe that the input of a JND model is the original image I_o and the output of a JND model is the JND map O_{JND}. The value of each pixel in the JND map O_{JND} represents the maximum tolerant intensity value variation of each pixel in the original image I_o. Different from the JND model, a VDP metric takes the original image I_o and the distorted image I_d as inputs. The output of a VDP metric is a probability map O_{VDP}. The value of each pixel in O_{VDP} represents the probability of detecting distortion at the corresponding pixel in the distorted image I_d. Overall, JND is a computational model of visual redundancy measured on a single original image while VDP is a distortion visibility metric measured on an image pair including an original image and its distorted counterpart. Although conceptually different, these two concepts are still relevant because they both attempt to model the same underlying mechanism of the visual system - detection and discrimination. In Section III-D, we will show how the JND map can be converted into a VDP-type metric and directly compared with existing VDP metrics.

The contributions of this work are as follows: 1) We propose a novel JND model for natural images from a top-down perspective without explicitly modeling and fusing different masking effects; 2) We transfer the problem of JND estimation into deriving a CPL counterpart of its original image and resort to the KLT theory for the first time to derive the critical point; 3) We conduct subjective experiments to determine the critical points of 500 images and find that the distribution of cumulative normalized KLT coefficient energy values at the critical points can be well characterized by a Weibull distribution; 4) We demonstrate the effectiveness of the proposed JND model in three different tasks including noise injection, image compression, and distortion visibility prediction.

The reminder of this paper is outlined as follows. Section II illustrates the proposed JND model including motivation and algorithm details. Section III presents the experimental results and comparisons with existing JND models. Finally, conclusions are drawn in section IV.

II. Methodology

A. Design Philosophy and Motivation

As stated, we turn to a top-down design philosophy for JND estimation. As illustrated in Fig. 1, most of the traditional JND profiles follow a bottom-up design philosophy. Given an input image, they start from visibility masking effect modeling of multiple factors including LA, CM, PC, CSF, etc. Afterwards, those masking effects are fused together to obtain a final JND estimation via linear or nonlinear combinations. However, this bottom-up design philosophy suffers from some inherent drawbacks. First, the overall visibility masking effect of the HVS can be related with more contributing factors beyond those have been considered in the existing works and it is also insufficiently accurate to formulate the masking effect even for a single specific contributing factor. Moreover, the used linear or nonlinear models for different masking effect fusion are also unable to characterize the complex interactions among different masking effects. To overcome such drawbacks, we propose a top-down design philosophy.

Considering the visibility limitation of HVS, it is believed that a CPL counterpart of the input image exists. The CPL image refers to a changed image whose change magnitude is just at the critical point beyond which this changed image will be either perceived as lossy or not the critical one (although perceptual lossless). Since we just cannot perceive the difference in the CPL, thus the difference map between the original image and the CPL image is just the JND map according to the definition of JND. Thus, the problem of JND estimation can be transfered into a CPL image estimation problem. Compared with directly estimating the JND map, it is much easier and intuitive to derive a CPL image. In this work, we resort to the KLT to obtain the CPL image.

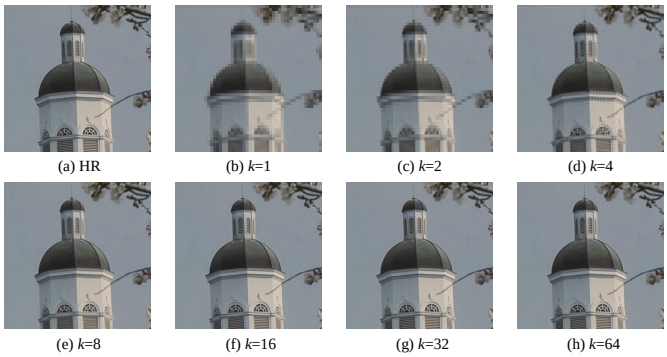
KLT is a signal-dependent linear transform. As a data-driven transform, it KLT has been applied in image coding [38] and image quality assessment [39, 40] with promising performance due to its excellent decorrelated performance.

The KLT domain and spatial domain can be converted from one to another without loss of information. Given an input image, we first transfer it from the spatial domain to the KLT domain. The KLT coefficients associated with different spectral components are responsible for visual information in different aspects. The former spectral components are responsible for macro-structures image information while the latter spectral components are responsible for micro-structures image information. To validate this point, Fig. 2 illustrates the image reconstruction process via inverse KLT, i.e., from KLT domain to spatial domain, with different numbers of spectral components. In this example, we set the total number of spectral components $K = 64$. Fig. 2(a) is the original image, Fig. 2(b)-(h) are the reconstructed images with the first k spectral components, where $k \in \{1, 2, 4, 8, 16, 32, 64\}$. As shown in Fig. 2(b), when only the 1-st spectral component is involved for reconstruction, almost all the macro-structures are recovered. However, many small textures and fine details are missing. As k increases, the small textures and fine details become richer and clearer. These observations well validate our previous statement that different spectral components are responsible for image information in different aspects. Generally, we have the insight that the front part of spectral components in the KLT coefficient matrix take charge of the reconstruction of image macro-structures such as the basic contour and main structures while the latter part of spectral components take charge of the reconstruction of image micro-structures such as the textures and fine details.

Obviously, there is a redundancy of visual information in an image. As shown in Fig. 2(f), when we take the first 16 spectral components to perform reconstruction via inverse KLT, the visual quality of the reconstructed image is almost the same with that of the original image, i.e., Fig. 2(a). The image reconstruction process via inverse KLT demonstrates that some micro-structure image information is perceptually redundant to HVS and directly discard the corresponding spectral components in the KLT coefficient matrix for reconstruction would not cause visible visual quality degradation of the reconstructed image. We wonder that if there exists a ciritical point (perceptual lossless threshold) L which satisfies the following property: when we take the first L spectral components to perform image reconstruction via inverse KLT, all the macro-structure image information and sufficient micro-structure image information are recovered so as to reconstruct the CPL image. Obviously, the accurate determination of the critical point L is a key step to the success of our JND model as different critical points L will yield different CPL images as well as different JND maps.

B. JND Estimation Based on CPL Image Prediction

1) Image Transfom Using KLT: KLT is a signal dependent linear transform, the kernels of which are derived by computing the principal components along eigen-directions of the autocorrelation matrix of the input data. Given an image $\mathbf{X}$ with size $M \times N$, a set of non-overlapping patches with size $\sqrt{K} \times \sqrt{K}$ are extracted. These image patches are vectorized and combined together to form a new matrix



(a) HR (b) k=1 (c) k=2 (d) k=4
(e) k=8 (f) k=16 (g) k=32 (h) k=64

Fig. 2: Reconstructed images with different numbers of spectral components in KLT. Zoon-in for best viewing.

$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_S] \in \mathbb{R}^{K \times S}$, where $\mathbf{x}_s \in \mathbb{R}^{K \times 1}, s = 1, 2, \cdots, S$ represents the s-th vectorized patch and S is the total number of image patches in $\mathbf{X}$. The covariance matrix of $\mathbf{X}$ is defined as follows

$$\mathbf{C} = \mathbb{E}[(\mathbf{x}_s - \bar{\mathbf{x}})(\mathbf{x}_s - \bar{\mathbf{x}})^{\mathrm{T}}] \tag{3}$$

$$= \frac{1}{S-1} \sum_{s=1}^{S} (\mathbf{x}_s - \bar{\mathbf{x}})(\mathbf{x}_s - \bar{\mathbf{x}})^{\mathrm{T}} \tag{4}$$

where $\bar{\mathbf{x}} = \frac{1}{S} \sum_{s=1}^{S} \mathbf{x}_s$ denotes the mean vector obtained by averaging each row of $\mathbf{X}$ and $\mathbf{C} \in \mathbb{R}^{K \times K}$. Then, the eigenvalues and eigenvectors of $\mathbf{C}$ are calculated via eigenvalue decomposition. The eigenvectors are arranged according to their corresponding eigenvalues in the descending order to form the KLT kernel $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_K] \in \mathbb{R}^{K \times K}$ where $\mathbf{p}_k \in \mathbb{R}^{K \times 1}$, $k = 1, 2, \cdots, K$ represents the k-th eigenvector. Using the KLT kernel $\mathbf{P}$, the KLT of $\mathbf{X}$ is expressed as follows:

$$\mathbf{Y} = \mathbf{P}^{\mathrm{T}} \mathbf{X} \tag{5}$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K]^{\mathrm{T}} \in \mathbb{R}^{K \times S}$ is the KLT coefficient matrix and $\mathbf{y}_k \in \mathbb{R}^{S \times 1}$, $k = 1, 2, \cdots, K$ refers to the k-th spectral component obtained by $\mathbf{y}_k = (\mathbf{p}_k)^{\mathrm{T}} \mathbf{X}$.

2) Convergence Property of KLT Coefficient Energy: After obtaining the KLT coefficient matrix $\mathbf{Y}$, we calculate the KLT coefficient energy E_k for spectral component y_k as follows:

$$E_k = \frac{1}{S} \sum_{s=1}^{S} \mathbf{Y}(k, s)^2, \quad k = 1, 2, \cdots, K. \tag{6}$$

In order to remove the influence of image content, we further calculate the normalized KLT coefficient energy p_k as follows:

$$p_k = \frac{E_k}{E_1 + E_2 + \cdots + E_K}, \quad k = 1, 2, \cdots, K. \tag{7}$$

Fig. 3(a) shows the normalized KLT coefficient energy distribution curve for the image in Fig. 2(a). Since p_1 is particularly large than others, we further plot another curve in Fig. 3(b) by excluding p_1. As shown in Fig. 3(b), as the spectral component index k increases, the KLT coefficient energy p_k first drops dramatically and later converges to be stable. When k is larger than 20, p_k becomes extremely small and gradually converges to zero. This phenomenon is consistent with the inverse KLT-based image reconstruction process illustrated

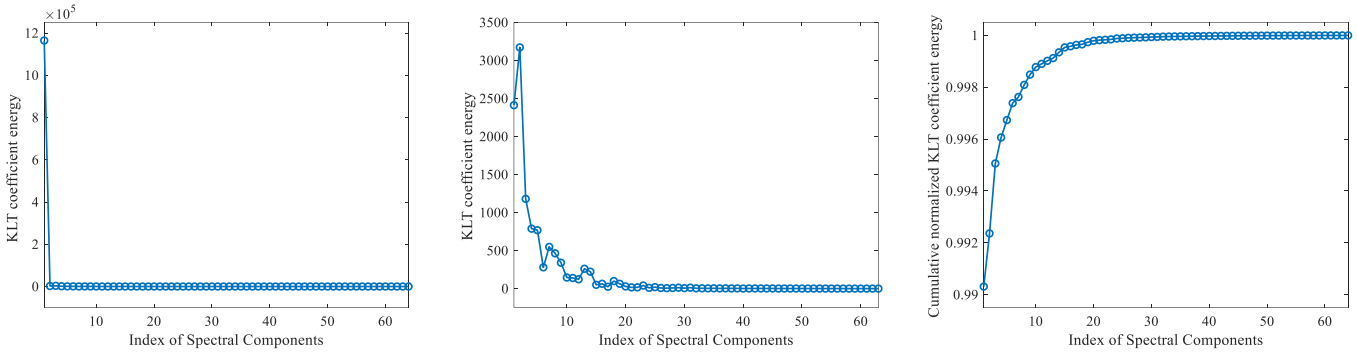


Fig. 3: Distribution curves of the normalized KLT coefficient energy and cumulative normalized KLT coefficient energy.

in Fig. 2. The former spectral components occupy much larger energies than the later ones so that they take charge of the macro-structure image information. The latter spectral components own relatively small energies such that they are in charge of the micro-structure image information.

Now, let us take a look at the normalized KLT coefficient energy curve from an cumulative perspective. The cumulative normalized KLT coefficient energy P_k is obtained as follows:

$$P_k = p_1 + p_2 + \cdots + p_k, \quad k = 1, 2, \cdots, K. \tag{8}$$

The cumulative distribution curve of the normalized KLT coefficient energy for the image in Fig. 2(a) has been shown in Fig. 3(c). Note that $P_1 = p_1$ and has been already close to 1 due to the particularly large energy of the first spectral component. As k increases, P_k monotonously increases. When $k > 20$, P_k will gradually converges to 1, indicating the recovered visual information gradually become saturated.

3) Critical Point Derivation: For images with different contents, although their cumulative distribution curves have the same convergence tendency, their critical points that lead to sufficient visual information may be different, as demonstrated in Fig. 4. Therefore, it is required to design an adaptive scheme to automatically determine the critical point for different images. In the following, we will illustrate how this issue is addressed with subjective studies and statistical analyses.

We select a total number of 500 high visual quality images from the DIV2K [41, 42] dataset. For each image, we apply the KLT-based image transform and inverse KLT-based image reconstruction with the first k spectral components, $k = 1, 2, \cdots, K$. Thus, we can obtain K reconstructed images for each original image. Then, we conduct a user study to subjectively determine the critical point $L \in \{1, 2, \cdots, K\}$ for each original image. Details of the user study are illustrated as follows. There are 60 participants (38 males and 22 females) in our subjective experiment. Each participant s^m is asked to compare each original image I_o and its corresponding K reconstructed versions $\{I_r^k\}$ one-by-one. During the subjective experiments, the image pairs are presented in a pre-defined order: $\{I_o, I_r^1\}$,$\{I_o, I_r^2\}$,$\cdots$,$\{I_o, I_r^K\}$. The image pairs are displayed on a DELL U2419HS monitor without resizing. The display is set to the sRGB color profile, the size of display is 23.8 inches (593.5mm×353.5mm), the resolution

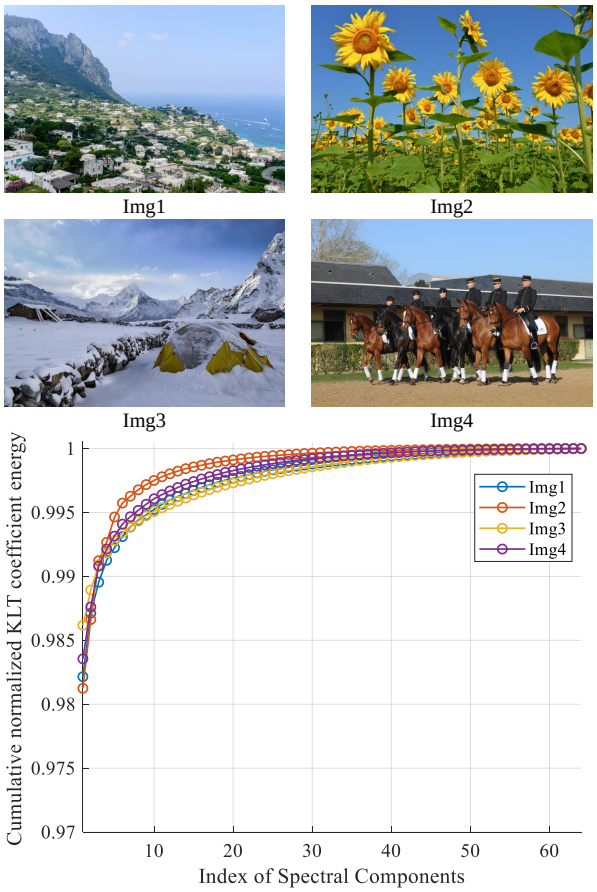


Fig. 4: Cumulative distribution curve of the normalized KLT coefficient energy.

is 1920×1080, the peak luminance is set to 220cd/m^2, the contrast is 1000:1, the viewing distance is 0.45m, and the pixels per visual degree (ppd) is 29.6266, approximately 30. All the subjective experiments are conducted in a dark room with dimmed lights. Each participant observes the presented image pairs carefully and conduct a binary judgment to answer the question whether the presented image pair has visible difference or not. Typically, the participants will observe obvious difference for the first several pairs and then gradually have difficulties in differentiating the later ones. For each

TABLE I: Data analysis on vote distribution of different images.

Image	P-value	$\mu_{\mathbf{L}_n}$	$\sigma_{\mathbf{L}_n}$	$[\mu_{\mathbf{L}_n} - 3\sigma_{\mathbf{L}_n}, \mu_{\mathbf{L}_n} + 3\sigma_{\mathbf{L}_n}]$	$\mu_{\mathbf{L}_n'}$	L_n	P_{L_n}
Img1	0.11917318	22.0000	4.6080	[8.1761, 35.8239]	21.7288	22	0.99787338
Img2	0.28491706	11.9667	3.6695	[0.9580, 22.9753]	11.9667	12	0.99802384
Img3	0.00020296	24.8361	3.8033	[13.4262,36.2459]	25.0833	26	0.99796027
Img4	0.09625943	18.2167	4.5829	[4.4679, 31.9654]	17.9661	18	0.99788544

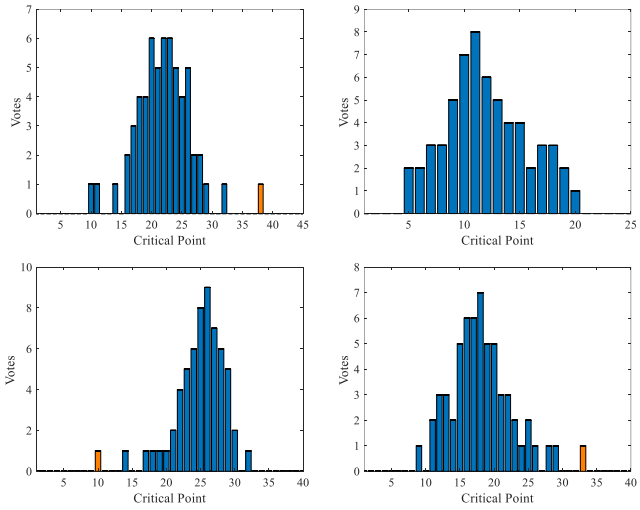


Fig. 5: Visualization of vote distributions from all participants.

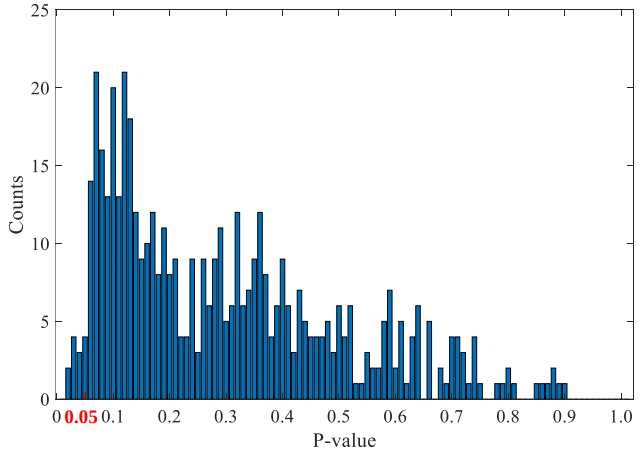


Fig. 6: Histogram of the P-values for all 500 images.

participant, the user study will stop once he/she cannot observe any visible difference for a certain image pair $\{I_o, I_r^k\}$ and we cosider this spectral component index k as the critical point given by the m-th participant s^m for the n-th original image I_o^n: $L_n^m = k$. The experiment then continues untill each participant have finished determining the critical points for all 500 images. To understand the distribution of subjective votes, we also take the four images shown in Fig. 4 as examples and plot their vote distributions from all participants in Fig. 5. We can see that the votes of critical points from all participants for each individual image approximate a Gaussian distribution.

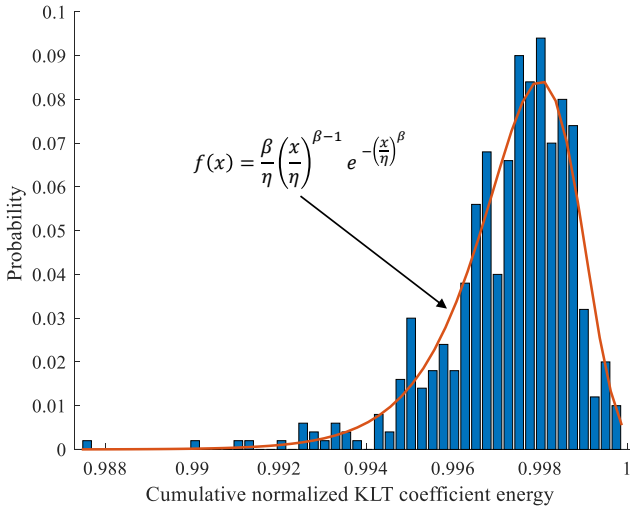


Fig. 7: Visualization of the distribution of cumulative normalized KLT coefficient energy over all 500 original images. The fitted parameters are $\beta = 894.16$ and $\eta = 0.998$.

Let us denote the critical points given by all participants for the n-th original image as $\mathbf{L}_n = [L_n^1, L_n^2, \cdots, L_n^{60}]$. For Gaussian-like distribution, we remove the outlier data points from $\mathbf{L}_n$ according to the well-known 3-σ criterion which assumes that a set of test data only contains random errors, calculate it to obtain standard deviation, and determine a range according to a certain probability of 99.7% [43]. It is considered that the error exceeds this interval is not a random error. That is, a specific element L_n^m will be excluded if it does not satisfy $\mu_{\mathbf{L}_n} - 3\sigma_{\mathbf{L}_n} \leq L_n^m \leq \mu_{\mathbf{L}_n} + 3\sigma_{\mathbf{L}_n}$ where $\mu_{\mathbf{L}_n}$ and $\sigma_{\mathbf{L}_n}$ denote the mean value and standard deviation of $\mathbf{L}_n$, respectively. The identified outliers have been marked in orange. After outlier removal, $\mathbf{L}_n$ will be updated to be $\mathbf{L}_n^{'}$. We further calculate the mean value $\mu_{\mathbf{L}_n'}$ and set the final critical point for the n-th original image as follows:

$$L_n = \lceil \mu_{\mathbf{L}_n'} \rceil, \tag{9}$$

where $\lceil \cdot \rceil$ represents the round up operation. A detailed data analysis on vote distribution of different images is summarized in Table I.

In order to understand whether the vote distributions for all 500 images follow the normal distribution or not, we further conduct a normality test. Here, the Shapiro-Wilk test (i.e., SW test) [44] is adopted. The significance level is set to α=0.05 [44]. If the P-value is less then α, it means the sample data

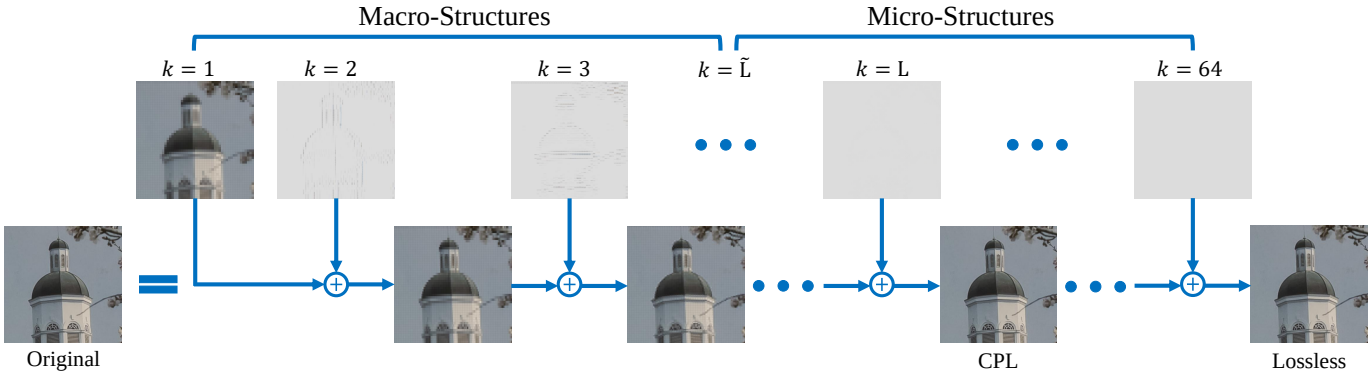


Fig. 8: Visualization of the progressive image reconstruction process via inverse KLT.

is significantly different from the normal distribution. Fig. 6 shows the histgram of all 500 P-values. It is observed that only 13 out of all P-values were less than 0.05, only accounting for 2.6%. This demonstrates that the vote distributions of most images (97.4%) conform to normal distribution. Thus, our strategy using the 3-σ criterion to filter the outliers is rational and appropriate.

Next, we calculate the corresponding cumulative KLT coefficient energy P_{L_n} for each original image with the obtained critical points $L_n \in \{1, 2, \cdots, K\}$. As a result, we can obtain a vector $\mathbf{P}_L = [P_{L_1}, P_{L_2}, \cdots, P_{L_{500}}]$ with each element representing the critical cumulative KLT coefficient energy of a specific original image. The calculated cumulative KLT coefficient energy values of the same four images are given in the last column in Table I. The distribution of cumulative normalized KLT coefficient energy over all 500 original images is also presented in Fig. 7. It is found that, although the critical points vary with different image contents, their corresponding cumulative KLT coefficient energy values tend to highly concentrate on the range of [0.99,1]. We can see from the figure that the distribution is right-skewed and can be well approximated by a Weibull distribution [45]. The Weibull distribution is a continuous probability distribution typically used for fitting both left- and right-skewed data. The widely used two-parameter Weibull distribution is expressed as follows:

$$f(x) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^{\beta}}, \beta > 0, \eta > 0, x > 0 \qquad (10)$$

where β and η are the shape parameter and scale parameter, respectively. The above distribution can be well fitted by the Weibull distribution with $\beta = 894.16$ and $\eta = 0.998$. The fitted Weibull distribution curve (the red curve in Fig. 7) can be considered as a statistical prior which directly shows the probability of each k (the index of spectral component) to be determined as the critical point.

Given a test image, the expected value of k is calculated to derive its corresponding critical point where the probabilities are determined by the fitted Weibull distribution function:

$$L = \left\lceil \frac{\sum_{k=1}^{K} k \cdot f(P_k)}{\sum_{k=1}^{K} f(P_k)} \right\rceil, \qquad (11)$$

where $\lceil \cdot \rceil$ represents the round up operation.

When we take the first L spectral components for image reconstruction via inverse KLT, it is expected to generate the CPL counterpart of the original image. How the CPL image can be reconstructed with the first L spectral components will be detailed in the following subsection.

4) CPL Image Reconstruction: To utilize the first L spectral components for image reconstruction via inverse KLT, we first define the corresponding reconstruction KLT coefficient matrix $\mathbf{Y}^{(L)}$ as follows:

$$\mathbf{Y}^{(L)} = \begin{bmatrix} y_{1,1} & \cdots & y_{L,1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{1,S} & \cdots & y_{L,S} & 0 & \cdots & 0 \end{bmatrix}^{\mathrm{T}} \qquad (12)$$

where $\mathbf{Y}^{(L)} \in \mathbb{R}^{K \times S}$, $k = 1, 2, \cdots, K$. Then, the image is reconstructed as follows:

$$\mathbf{X}^{(L)} = \mathbf{P}\mathbf{Y}^{(L)} \qquad (13)$$

where $\mathbf{X}^{(L)}$ represents the reconstructed image by only considering the first k spectral components. Since L is the estimated critical point, $\mathbf{X}^{(L)}$ can be considered as the CPL image. Fig. 8 illustrates the progressive image reconstruction process. In our experiment, we set $K = 64$. The images shown in the top row are the reconstructed results using only each individual single spectral component while the images shown in the bottom row are the reconstructed results using all previous spectral components. It is obvious that only using the first spectral component for reconstruction could recover most macro-structures of the original image. When more spectral components are involved, the image is progressively reconstructed with richer and finer details. Suppose $\tilde{L}$ is the boundary between macro- and micro-structures. Using the former $\tilde{L}$ spectral components will successfully reconstruct all the macro-structures and the latter $(64 - \tilde{L})$ spectral components will mainly responsible for the reconstruction of micro-structures. For the critical point L we have estimated, it means that we at least need to use the former L spectral components to reconstruct all the macro-structures and sufficient micro-strucutres required for a perceptual lossless image. Finally, if all the spectral components are involved, the original image can be completely reconstructed.

In Fig. 9, we present our predicted CPL images of the four original images. As we can see, the CPL images (second

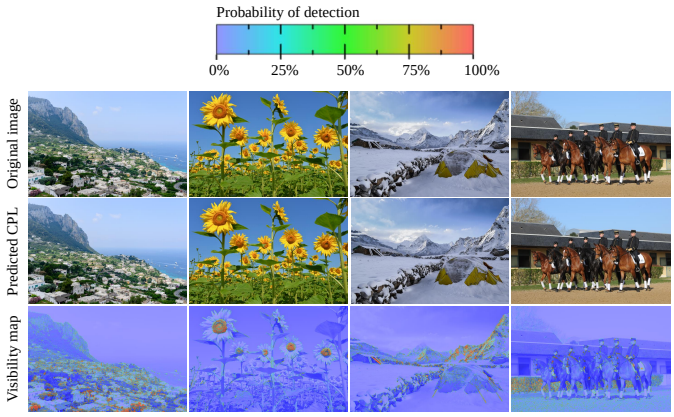


Fig. 9: Visualization of the predicted CPL images and the distortion visibility maps estimated by HDR-VDP2.2 [46].

row) are quite similar to the original ones, without obvious visible distortions. We also use the HDR-VDP2.2 metric [46] to calculate the distortion visibility maps, as shown in the third row. It is observed that most regions in the visibility maps are blue, implying that it is hard to perceive distortions from these CPL images. These results show that our determined CPL images are close to their original counterparts from the perspective of human perception.

5) JND Map Estimation: Once the CPL image $\mathbf{X}^{(L)}$ is obtained, we compute the difference map between the original image $\mathbf{X}$ and $\mathbf{X}^{(L)}$ as the final JND map $\mathbf{M}$:

$$\mathbf{M}(i,j) = |\mathbf{X}(i,j) - \mathbf{X}^{(L)}(i,j)|, \tag{14}$$

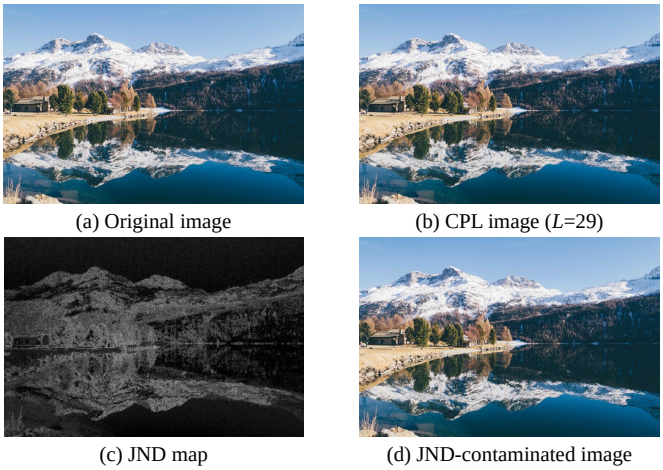
where $|\cdot|$ denotes the absolute value operator, (i,j) are the pixel coordinates in spatial domain.

III. Experiments

In this section, the proposed JND model is compared with the existing JND models to demonstrate its accuracy in direct prediction of JNDs and efficiency in eatimation of scene complexity and perceptual redundancy.

A. Performance Comparison on Direct JND Prediction

In order to directly show the model ability to predict JND, we use the dataset in [34] for validation. This dataset provides 20 reference images (i0webp-i19webp) along with their corresponding VLT data under different viewing conditions. According to the viewing conditions setting in our study, we select the VLT data measured under the viewing condition of 220$\mathrm{cd/m^2}$ peak luminance and 30 ppd. For each reference image in the dataset, we find the visually lossless version based on the provided VLT data. Thus, the difference map between the reference image and the visually lossless version is considered as the ground-truth JND map. Then, for each reference image, we also predict its JND maps by different JND models. Then, we conduct a normalization process on the predicted JND map and the ground-truth JND map, respectively. Specifically, for each pixel in the JND map, we divide it by its maximum value in that map and all pixel values will fall into the range of [0,1]. After normalization, we calculate the RMSE value between the predicted JND map and the ground-truth JND map. A smaller RMSE value means a higher similarity between the predicted JND map and the ground-truth JND map, thus directly showing a better ability of a specific model to predict JND. The experimental results are listed in Table II where the best performer (smallest RMSE) for each image has been bolded. We can see that our proposed JND model owns the smallest RMSE values on all test images, demonstrating the best ability to predict JND.



(a) Original image (b) CPL image (L=29)

(c) JND map (d) JND-contaminated image

Fig. 10: An example of noise contaminated image guided by our JND map. The PSNR is 26dB.

B. JND-Guided Noise Injection

According to previous works [23–27], the performance of a JND model can be measured by the capability of hiding noise to some extent. Specifically, we can get the JND-contaminated image by injecting random noise into the image with the guidance of the estimated JND map as follows:

$$\widetilde{\mathbf{X}}(i,j) = \mathbf{X}(i,j) + \theta \cdot \mathbf{N}(i,j) \cdot \mathbf{M}(i,j), \tag{15}$$

where $\widetilde{\mathbf{X}}$ is the JND-contaminated image, $\mathbf{N}$ denotes the bipolar random noise of ± 1, and θ is a noise energy regulating factor. By adjusting the value of θ, the amount of noise injected into the image can be well controlled. In other words, we can inject almost the same amount of noise into the test image (almost the same PSNR) by adjusting the value of θ for different JND maps.

In Fig. 10, we give an example to show the results of nosie contaminated image guided by our proposed JND map. Obviously, it is difficult to perceive the visual quality difference between the original image (i.e., Fig. 10(a)) and the contaminated image (i.e., Fig. 10(d)). That is, although the PSNR value is only 26dB, the injected noise are not visible. From the JND map shown in Fig. 10(c), we observe that many details in the complex textured regions are redundant to the HVS. Using such a JND map as a guidance for noise injection, the contaminated image will be perceived with high quality since most of the noises are encouraged to be added in those highly redundant area that is not sensitive to the HVS.

TABLE II: RMSE values between the predicted JND maps and the ground-truth JND maps. A smaller RMSE value means a better performance.

Image	Yang2005 [21]	Zhang2005 [47]	Wu2013 [23]	Wu2017 [24]	Jakhetiya2018 [25]	Chen2020 [26]	Shen2021 [27]	Proposed
i0webp	0.3865	0.2408	0.1521	0.1918	0.1822	0.2611	0.2476	**0.1032**
i1webp	0.5284	0.2614	0.1321	0.1809	0.2011	0.3749	0.2530	**0.0972**
i2webp	0.3714	0.2625	0.1580	0.2849	0.2099	0.2234	0.2197	**0.1447**
i3webp	0.2255	0.1956	0.1063	0.1199	0.1072	0.2522	0.2024	**0.0816**
i4webp	0.5479	0.2567	0.1580	0.1992	0.1730	0.3412	0.2667	**0.0795**
i5webp	0.6412	0.2813	0.1741	0.2422	0.2334	0.4649	0.2201	**0.0982**
i6webp	0.3689	0.2409	0.1264	0.1633	0.1651	0.2815	0.2070	**0.1202**
i7webp	0.3151	0.1995	0.1207	0.1852	0.1842	0.2350	0.2540	**0.1017**
i8webp	0.4960	0.2535	0.1381	0.1774	0.1923	0.2742	0.2853	**0.1075**
i9webp	0.3420	0.3196	0.1568	0.1665	0.2021	0.2728	0.2114	**0.1228**
i10jpeg	0.3290	0.2383	0.1778	0.1841	0.2516	0.1923	0.2066	**0.1172**
i11jpeg	0.3526	0.1912	0.0899	0.1684	0.1151	0.2477	0.2147	**0.0743**
i12jpeg	0.3538	0.2039	0.1636	0.3025	0.2360	0.2666	0.1890	**0.1051**
i13jpeg	0.2423	0.2049	0.0985	0.1099	0.1020	0.2010	0.1440	**0.0502**
i14jpeg	0.3078	0.2822	0.0960	0.1293	0.1273	0.2371	0.1555	**0.0713**
i15jpeg	0.2280	0.2208	0.0906	0.1216	0.1094	0.2786	0.2345	**0.0378**
i16jpeg	0.3401	0.2311	0.1244	0.1638	0.1526	0.2407	0.2284	**0.0949**
i17jpeg	0.3521	0.2781	0.1190	0.1646	0.1689	0.3218	0.1889	**0.0790**
i18jpeg	0.3990	0.2285	0.1049	0.1541	0.1706	0.3074	0.2524	**0.0779**
i19jpeg	0.4718	0.2697	0.0901	0.1268	0.1693	0.3359	0.2400	**0.0678**
Average	0.3800	0.2430	0.1289	0.1768	0.1727	0.2805	0.2211	**0.0916**

TABLE III: Performance comparison of different JND models on noise injection. The HDR-VDP score is based on HDR-VDP2.2 [46].

Image	Yang2005 [21]		Zhang2005 [47]		Wu2013 [23]		Wu2017 [24]		Jakhetiya2018 [25]		Chen2020 [26]		Shen2021 [27]		Proposed	
	HDR-VDP	MOS	HDR-VDP	MOS	HDR-VDP	MOS	HDR-VDP	MOS	HDR-VDP	MOS	HDR-VDP	MOS	HDR-VDP	MOS	HDR-VDP	MOS
I01	55.8219	-0.8333	54.4010	-0.6333	56.6592	-0.6000	58.3114	-0.5667	58.7397	-0.5778	55.8872	-0.3889	**58.9145**	-0.6333	58.1479	**-0.2313**
I02	56.0151	-0.6444	54.4634	-0.7222	58.0067	-0.4667	59.3012	-0.3889	59.5552	-0.3778	53.7419	-0.6222	54.5701	-0.5778	**59.8919**	**-0.1938**
I03	54.2115	-0.6556	53.4778	-0.7333	56.1152	-0.6000	57.8486	-0.5333	56.4067	-0.5889	54.3552	-0.6778	53.7209	-0.5889	**58.2178**	**-0.2125**
I04	53.7742	-0.5556	52.1998	-0.7111	55.0556	-0.5111	57.0854	-0.3556	55.7968	-0.4111	52.5509	-0.8222	**58.3130**	-0.3778	56.9190	**-0.1938**
I05	52.3229	-0.7111	50.4194	-0.6222	52.1538	-0.5778	54.2264	-0.4222	52.9179	-0.4444	52.1209	-0.5889	**57.3850**	-0.3778	57.2651	**-0.2563**
I06	52.3135	-0.6222	52.5651	-0.5444	54.9231	-0.5000	56.2502	-0.3556	56.4490	-0.3889	52.3666	-0.7222	**58.6321**	-0.3444	57.0403	**-0.2438**
I07	56.2070	-0.6333	54.7842	-0.6333	56.3356	-0.5222	**60.3065**	-0.4111	58.4491	-0.3556	52.5698	-0.7111	54.0611	-0.6333	58.7755	**-0.2563**
I08	53.9325	-0.4778	52.7390	-0.5778	55.3771	-0.5222	57.7465	-0.2889	56.5790	-0.5667	52.6583	-0.7222	55.0882	-0.3778	**58.6738**	**-0.2375**
I09	54.2341	-0.5667	52.5777	-0.6444	54.6901	-0.5333	56.8661	-0.3778	54.1190	-0.4889	52.5546	-0.7444	52.5290	-0.6333	**56.9770**	**-0.2688**
I10	55.1364	-0.6000	53.0932	-0.6667	55.3694	-0.5222	**57.1095**	-0.4111	56.4227	-0.3444	54.8467	-0.5333	54.4159	-0.3556	56.5825	**-0.2125**
I11	53.0243	-0.6000	51.6517	-0.6000	54.2494	-0.4778	56.0900	-0.3000	53.6028	-0.4000	51.9667	-0.7222	52.5632	-0.6333	**57.0014**	**-0.2438**
I12	53.4857	-0.4333	51.0445	-0.4000	53.4636	-0.3444	54.4890	-0.3778	54.4000	-0.3000	53.1855	**-0.2778**	51.6240	-0.3556	**55.5404**	-0.3438
I13	58.0653	-0.5556	56.2722	-0.5778	58.6279	-0.4778	**61.3581**	-0.3667	60.1869	-0.4000	52.1352	-0.6778	55.7032	-0.6778	60.3400	**-0.2688**
I14	55.5298	-0.3667	54.5143	-0.4667	57.3347	-0.4111	59.8811	-0.3000	56.2978	-0.4889	51.9380	-0.7889	57.5791	-0.6444	**60.5972**	**-0.2375**
I15	56.0810	-0.4111	53.1295	-0.4889	56.3243	-0.4222	**57.4761**	-0.3000	56.4456	-0.4556	54.0723	-0.7444	56.7220	-0.5778	57.3237	**-0.2125**
I16	50.9370	-0.3333	50.3087	-0.5333	51.8376	-0.5000	53.2925	-0.3444	52.6953	-0.3778	51.2403	-0.6333	51.1785	-0.7333	**54.9153**	**-0.2875**
I17	53.1796	-0.4222	53.4795	-0.6222	55.1712	-0.4667	**57.0430**	-0.3111	55.8383	-0.4000	54.0341	-0.6778	52.4972	-0.3556	56.6582	**-0.2750**
I18	52.2712	-0.4222	52.4559	-0.5444	54.4790	-0.3778	58.0745	-0.2333	56.3035	-0.3000	52.0321	-0.6333	51.3789	-0.5778	**59.1816**	**-0.2625**
I19	52.1117	-0.3778	52.1802	-0.5889	54.3710	-0.4333	57.6392	-0.3222	55.7807	-0.3778	52.1641	-0.7111	57.1253	-0.5000	**58.1924**	**-0.2375**
I20	53.6226	-0.3778	51.4270	-0.5444	54.2392	-0.4667	56.1542	-0.2889	54.4725	-0.3667	52.1627	-0.6333	49.4429	-0.6333	**56.7589**	**-0.2438**
Average	54.1139	-0.5300	52.8592	-0.5928	55.2392	-0.4867	57.3275	-0.3628	56.0729	-0.4206	52.9292	-0.6517	54.6722	-0.5294	**57.7500**	**-0.2459**

1) Performance Comparison on JND-Guided Noise Injection: Apparently, with the guidance of a more accurate JND map, the JND-contaminated image (obtained by Eq. (13)) with same noise level should have better visual quality. As stated, for different JND models, we can adjust the value of θ to ensure that almost the same amount of noise is injected. Then, the performances of differenta JND model can be compared objectively and subjectively. Specifically, objective evaluation is performed based on the HDR-VDP2.2 metric [46] which is known as a popular distortion visibility detection metric for both high-dynamic range and standard images. A higher HDR-VDP score indicates less visible distortions. Subjective

Fig. 11: Twenty images used for testing in the experiments.

evaluation is performed by humans to obtain MOS score. Another 30 participants are invited to participate the subjective experiments. Each participant is asked to assign an opinion score in the range $[-1, 0]$ to a specific image pair including one original image and one JND-contaminated image. A score equals to 0 means the visual quality of the JND-contaminated image is equal to that of the original image while a score equal to -1 means the visual quality of the JND contaminated image is dramatically worse than that of the original image. As a result, for each contaminated image, 30 scores are collected from all participants. After removing the outlier, the mean value of the retained opinion scores is calculated as the MOS. Overall, a more accurate JND model will result in higher HDR-VDP score and MOS than the competitors.

We select another 20 high-quality images from the DIV2K [41, 42] dataset for testing. These images are shown in Fig. 11. The proposed JND model is compared with seven existing JND models including Yang2005 [21], Zhang2005 [47], Wu2013 [23], Wu2017 [24], Jakhetiya2018 [25], Chen2020 [26], and Shen2021 [27]. Table III provides the performance results of different JND models in terms of HDR-VDP score and MOS at the noise level of PSNR=26dB (comparisons under other different PSNR settings are also conducted and will be reported later in this section). It can be seen that our proposed JND model delivers higher HDR-VDP scores on 11 out of 20 test images and the highest HDR-VDP score in average when considering all 20 images. Followed by our proposed JND model, Wu2017 and Shen2021 take the first place for 5 and 4 times, respectively. In terms of MOS, we achieve the highest MOS values for almost all the images except the image 'I12' on which Chen2020 is the best. These results demonstrate the superiority of our proposed JND model against others in allocating noise at the noise level of PSNR=26dB (comparisons under other different PSNR settings are also conducted and will be reported later in this section). Note that we inject noise to the original images at the noise level of PSNR=26dB to obtain the JND-guided noise contaminated images. Generally, 26dB is a relatively large noise intensity which will probably break the transparency, and thus it is inevitable to induce visible distortions on the JND-guided noise contaminated image to some degree.

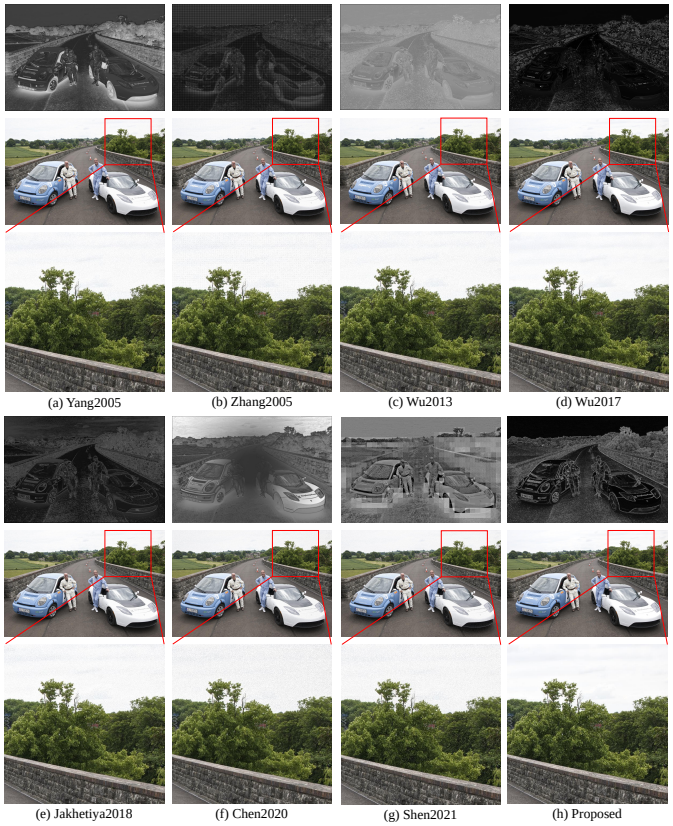


Fig. 12: Visual comparison of noise-contaminated images generated by using different JND models as guidance. Zoom-in for best viewing.

In order to make a more clear comparison among these JND models, Fig. 12 gives a visual example of the JND maps and the noise contaminated images. Images in the first row are the estimated JND maps. Images in the second row are the JND-guided noise-injected images generated according to Eq. (12). Images in the third row are the enlarged versions of sub-regions cropped from the noise-injected images in the second row. We can observe that the noise-contaminated image guided by our proposed JND model is the 'cleanest' one among all the compared ones. It seems that Fig. 12(h) is perceptually the same with the original image as it seems to have no visible noise. Among the competitors, Fig. 12(d) also has achieved satisfactory visual quality while it is still worse than Fig. 12(h). With careful observations, one can still perceive visible noise in the sky area of Fig. 12(d).

By taking a closer look at Fig. 12(h), we find that the injected noises are hidden in those textured areas such as the walls and leaves. However, it is hard to be perceived by the HVS at the first glance. As has been reported in [24], the HVS is highly adapted to extracting the repeated patterns for visual content representation and it is hard to perceive the noise which is injected into the textured areas with high pattern complexity. By contrast, it is much easier for the HVS to perceive the noise in those relatively smooth areas. Compared with the smooth areas, textured areas tend to contain much more details (e.g., micro-structures) and are

TABLE IV: Performance comparison on maximum tolerable noise level measured by PSNR. A smaller PSNR value means a better performance.

Image	Yang2005 [21]	Zhang2005 [47]	Wu2013 [23]	Wu2017 [24]	Jakhetiya2018 [25]	Chen2020 [26]	Shen2021 [27]	Proposed
I01	36.3267	37.0200	36.7256	36.5450	36.7110	37.1360	38.6131	**35.5836**
I02	31.7161	30.2181	30.4085	29.9859	**28.8173**	36.7045	31.4304	30.9822
I03	34.5942	37.6568	34.7696	36.8901	36.7906	40.9209	**33.4430**	34.4538
I04	32.0325	34.4976	31.7433	31.0419	33.2436	38.6290	36.8489	**30.9284**
I05	36.1455	34.9346	34.9324	34.5491	35.5152	39.4872	36.2114	**34.3448**
I06	34.1929	34.5028	33.0136	32.5747	33.8537	36.0527	36.8339	**32.3934**
I07	32.2634	32.0573	**31.0054**	31.7650	31.3990	35.1954	32.7384	31.4912
I08	33.9132	34.0813	31.9071	31.3348	35.5463	36.7112	43.1402	**31.1622**
I09	37.8955	33.9826	**32.5214**	38.6528	37.9368	38.2275	42.7123	36.4163
I10	32.4285	32.0577	32.4657	31.8867	31.9063	33.4762	33.7650	**30.7900**
I11	38.1032	37.5925	32.7340	**32.6700**	34.9148	37.2067	39.7867	35.8269
I12	35.7780	37.4056	36.5880	**35.4858**	35.9974	38.0843	37.6373	36.3014
I13	32.6346	32.6721	32.8009	32.5361	32.1162	36.2254	**31.2969**	31.8158
I14	31.8024	34.7035	32.1759	31.5463	34.2748	37.6211	37.7132	**30.8307**
I15	32.3944	34.3741	**32.0731**	33.0717	33.4020	34.7803	40.2603	33.4000
I16	36.9723	37.1364	**35.7520**	36.0867	37.7459	37.9674	41.6559	36.3747
I17	32.0907	40.6499	31.0099	31.0817	36.3353	34.4205	41.9413	**30.4702**
I18	31.4816	33.1370	33.0206	30.8787	36.5084	35.1940	35.7879	**30.4573**
I19	32.0469	30.8358	31.5857	30.8322	32.3963	34.5194	35.4254	**30.2432**
I20	32.9795	34.1478	33.1078	32.7811	33.9243	33.9742	41.6431	**32.3365**
Average	33.8896	34.6832	33.0170	33.1098	34.4668	36.6267	37.4442	**32.8301**

likely to be redundant. As we have demonstrated before, the difference map between the original image and the derived CPL image can well resemble the redundant micro-structure image information to the HVS. Thus, our proposed JND model using such difference map as the JND map will encourage to hide the noise in those highly redundant areas. Therefore, the injected noises will not be easily visible by the HVS. What's more, instead of modeling and aggregating the masking effects of diverse contributing factors in isolation, our proposed JND model dedicates to deriving a CPL image to best exploit the potential perceptual redundancies exist in the original image from a top-down perspective. Thus, our proposed JND model would be able to implicitly characterize the influence of more potential factors beyond those have been considered previously. As a result, our proposed JND model can finally generate a perceptually better noise-contaminated image.

Note that comparing the performance under PSNR=26dB only may be unfair to certain JND models. Therefore, we also conduct experiments under more PSNR settings. Specifically, we set PSNR=$\{22, 23, 24, 25, 26, 27, 28, 29, 30\}$ and for each PSNR setting we generate JND-contaminated images of I01-I20 by using different JND models as guidance. We also apply the HDR-VDP metric to compute the objective score of each JND-contaminated image, and we take the averaged HDR-VDP score as the indicator of JND model performance. Finally, we plot the curve of average HDR-VDP score versus PSNR value, as shown in Fig. 13. We can see that the curve of our proposed JND model is among the top, which indicates our JND model always achieves the highest HDR-VDP score at each PSNR value.

It should be emphasized that, although we have compared different JND models on noise injection with different noise levels, such experiments do not directly reflect the capability of predicting JND. The experimental setups for directly comparing the capability of predicting JND still deserve more rigorous treatments.

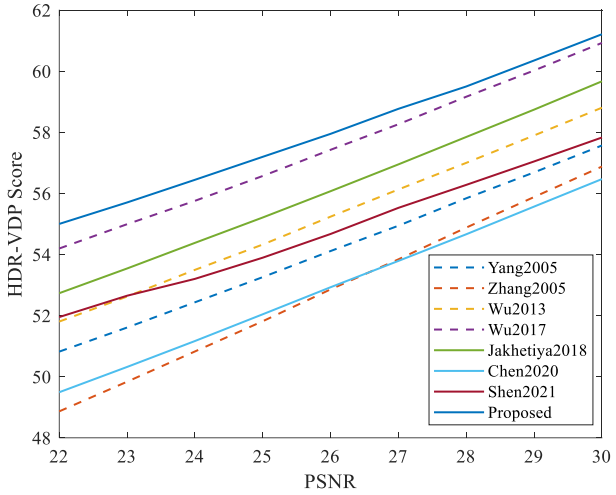
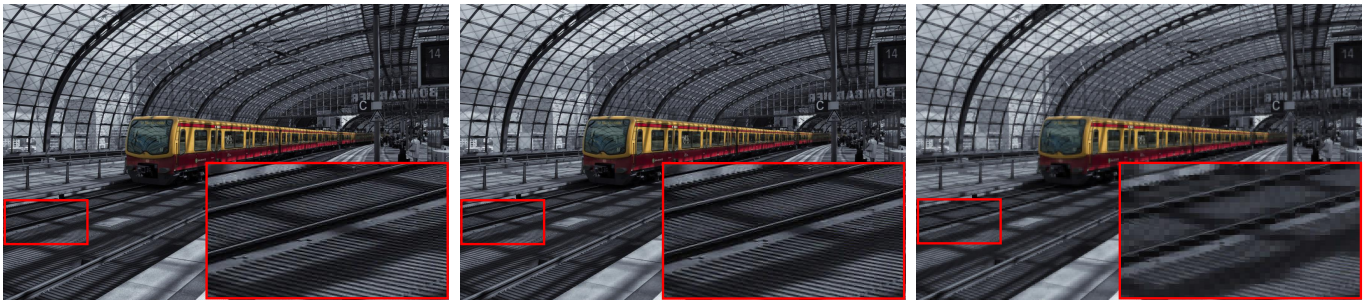


Fig. 13: Curve of average HDR-VDP score versus PSNR.

2) Performance Comparison on Maximum Tolerable Noise Level: The above experiments are conducted to compare the visual quality of different JND-guided noise contaminated images under the same PSNR level. While the results have demonstrated promising performance of our proposed JND model, it is still necessary to understand the maximum tolerable noise level by different JND models. A better JND model should tolerate more noise while keeping the quality of the noise-contaminated image perceptually unchanged. In our experiment, we use PSNR as a measure of the noise level and



(a) JPEG (b) JND (Proposed) + JPEG (c) JND (Wu2017) + JPEG

Fig. 14: Visual quality comparison of JND-guided image compression. (a) Direct JPEG compression result; (b) Our proposed JND model-guided JPEG compression result; (c) Wu2017 JND model-guided JPEG compression result.

TABLE V: Performance comparison on JND-guided JPEG compression (QF=1). A higher value of G means a better performance.

Image	JND (Proposed) + JPEG			JND (Wu2017) + JPEG		
	$\Delta Bitrate$	$\Delta PSNR$	G	$\Delta Bitrate$	$\Delta PSNR$	G
I01	9.07%	2.44%	3.7141	47.12%	17.46%	2.6988
I02	4.83%	1.08%	4.4883	39.64%	29.68%	1.3357
I03	23.24%	4.86%	4.7857	42.74%	11.30%	3.7819
I04	15.65%	3.55%	4.4047	43.47%	13.89%	3.1287
I05	23.83%	4.99%	4.7740	43.89%	12.51%	3.5099
I06	10.05%	2.19%	4.5872	38.10%	12.62%	3.0183
I07	11.44%	3.19%	3.5839	38.60%	15.07%	2.5610
I08	13.62%	4.19%	3.2484	44.82%	17.93%	2.4996
I09	13.43%	4.93%	2.7213	35.16%	15.98%	2.2006
I10	24.45%	5.30%	4.6149	46.94%	14.29%	3.2859
I11	14.66%	5.05%	2.9033	37.34%	16.35%	2.2839
I12	10.25%	3.01%	3.3996	44.10%	20.35%	2.1671
I13	13.33%	3.58%	3.7282	43.88%	14.37%	3.0536
I14	25.92%	4.61%	5.6230	51.00%	11.93%	4.2754
I15	18.70%	3.24%	5.7735	45.15%	13.19%	3.4214
I16	15.82%	4.50%	3.5121	36.45%	15.75%	2.3143
I17	24.62%	5.14%	4.7868	46.18%	12.35%	3.7384
I18	16.80%	3.32%	5.0552	45.34%	11.78%	3.8485
I19	21.54%	3.25%	6.6243	47.82%	8.98%	5.3267
I20	19.65%	4.89%	4.0204	39.94%	12.24%	3.2626
Average	16.54%	3.87%	4.2793	42.88%	14.90%	2.8779

a lower PSNR value actually indicates a better performance of a JND model. The results are listed in Table IV. As we can see, the proposed JND model can tolerable more noise on 11 out of 20 images and also has the lowest PSNR in average when considering all 20 images, which again validates the superiority of our proposed JND model.

C. JND-Guided Image Compression

Since the JND map implies the visibility limitation of the HVS. Thus, it is often employed in image compression to improve compression efficiency. Generally, the smoothing operation can reduce the signal variance, which makes image compression easier. However, blindly smoothing operations will always jeopardize image quality. Thus, we can use the JND map to guide the smoothing operation as a preprocess step before JPEG compression for perceptual redundancy reduction. As will be illustrated later, the visual quality of the compressed image guided by our proposed JND model will not be affected too much while saving considerable coding bits. Specifically, given an input image $\mathbf{T}$ and its corresponding JND map $\mathbf{T}_{\mathrm{M}}$, the JND-guided image smoothing operation is described as follows:

$$\widetilde{\mathbf{T}}(i,j) = \begin{cases} \mathbf{T}(i,j) + \mathbf{T}_{\mathrm{M}}(i,j), & \mathbf{T}(i,j) - \overline{\mathbf{T}}_{\mathrm{P}} < -\mathbf{T}_{\mathrm{M}}(i,j) \\ \mathbf{T}(i,j) - \mathbf{T}_{\mathrm{M}}(i,j), & \mathbf{T}(i,j) - \overline{\mathbf{T}}_{\mathrm{P}} > \mathbf{T}_{\mathrm{M}}(i,j) \\ \overline{\mathbf{T}}_{\mathrm{P}}, & else \end{cases} \tag{16}$$

where $\overline{\mathbf{T}}_{\mathrm{P}}$ denotes the mean value of the block that $\mathbf{T}(i,j)$ belongs to during the compression process (e.g., the divided 8×8 blocks that $\mathbf{T}(i,j)$ located at during JPEG compression).

With the above JND-guided image smoothing operation, the visual redundancy of the input image will be reduced to facilitate compression. One visual example is shown in Fig. 14 where we perform JPEG compression of the original image "I01" at QF=1 in two different ways: direct JPEG compression and JND-guided JPEG compression. Fig. 14(a) is the result obtained by direct JPEG compression. Fig. 14(b) is the result obtained by first preprocessing the original image with the JND-guided image smoothing operation and then performing JPEG compression. It can be observed that, though less bit rate is required (0.0684 bpp for Fig. 14(a) and 0.0622 bpp for Fig. 14(b)), the visual quality of Fig. 14(b) is almost equal to that of Fig. 14(a). As a comparison, we also provide the result obtained by using Wu2017 JND model as the guidance under the same QF setting in JPEG compression. As shown in Fig. 14(c), though the bit rate of Fig. 14(c) is 0.0362 bpp which is less than Fig. 14(c), we can easily perceive obvious compression artifact in this result.

We also notice that in comparison with Fig. 14(c), Fig. 14(b) achieves higher visual quality at the cost of higher bit rate. Therefore, an important issue is how to fairly compare the overall performance of different JND-guided JPEG compression results by jointly taking visual quality and bit rate into account with a single criteria. Keeping this in mind, we define the gain $G = \frac{\Delta Bitrate}{\Delta PSNR}$ as the criteria, where $\Delta Bitrate$ denotes the saving of bit rate and $\Delta PSNR$ denotes

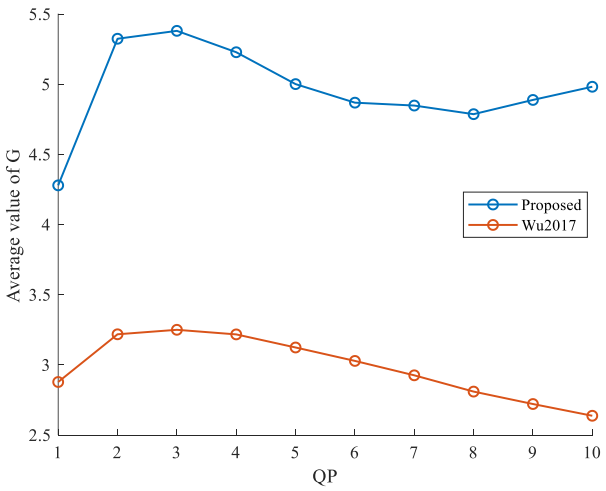


Fig. 15: The averge value of G with different QF settings in JPEG compression.

the reduction of PSNR value. Specifically, these two measures are calculated as follows:

$$\Delta Bitrate = \frac{Bitrate_{ori} - Bitrate_{jnd}}{Bitrate_{ori}} \times 100\% \quad (17)$$

and

$$\Delta PSNR = \frac{PSNR_{ori} - PSNR_{jnd}}{PSNR_{ori}} \times 100\%, \quad (18)$$

where $Bitrate_{ori}$ and $PSNR_{ori}$ represent the bit rate and PSNR value of direct JPEG compressed image, respectively, $Bitrate_{jnd}$ and $PSNR_{jnd}$ represent the bit rate and PSNR value of JND-guided JPEG compressed image, respectively. According this definition, it is reasonable to say a higher value of G indicates a better performance of a JND model for guiding the JPEG compression. We thus evaluate the performance of JND-guided JPEG compression using G and show the results in Table V. As shown, our proposed JND model-guided JPEG compression is able to save 16.54% of bitrates in average while Wu2017's JND model-guided JPEG compression can save 42.88% of bitrates in average. However, our proposed JND model-guided JPEG compression only reduces 3.87% of PSNR value in average while Wu2017's JND model-guided JPEG compression will reduce 14.90% of PSNR value in average. By taking PSNR reduction and bitrate saving into account simultaneously with a single criteria G, our proposed JND model-guided JPEG compression achieves much larger values of G for the majority images (except "I11") and also a much larger average value of G when considering all the images together.

We further set different QP values and draw the curves of the average value of G with different QP values, as depicted in Fig. 15. Our proposed JND model-guided JPEG compression consistently owns higher average value of G than that of Wu2017's by a large margin at each QP value. This further demonstrates the superiority of our proposed JND model for improving the efficiency of JPEG compression. Based on these experimental results, we can savely conclude that: 1) in comparison with the direct JPEG compression, the JPEG compression guided by our proposed JND model could further save moderate bitrates while without reducing the visual quality too much; 2) in comparison with Wu2017's JND model, JPEG compression guided by our proposed JND model could achieve a better balance between visual quality reduction and bit rate saving.

TABLE VI: RMSE between the distortion visibility maps predicted by VDP metrics and ground-truth maps marked by humans. JND2VDP represents the metric which is converted from our proposed JND model according to Eq. (19).

Subset	HDRVDP 2.0	HDRVDP 2.1	HDRVDP 2.2.1	HDRVDP 3.0.6	JND2VDP
aliasing	0.4450	**0.2012**	0.2046	**0.1315**	0.2344
cgibr	0.7790	**0.3714**	0.3773	**0.1706**	0.4467
compression	0.6958	0.5966	0.5950	**0.3442**	**0.5476**
deghosting	0.5912	**0.3488**	0.3495	**0.3339**	0.6090
downsampling	0.1524	**0.0846**	0.0849	**0.0732**	0.0861
ibr	0.7415	**0.2957**	0.3010	**0.1211**	0.4130
mixed	0.4495	**0.2243**	0.2289	**0.1236**	0.2638
perceptionpatterns	0.6916	0.4033	0.4068	**0.2952**	**0.3430**
peterpanning	0.4643	0.3717	0.3746	**0.1627**	**0.1856**
shadowacne	0.4787	0.3316	0.3347	**0.1630**	**0.2011**
tid2013	0.5060	**0.4380**	**0.4400**	0.4760	0.5038
zfighting	0.4664	0.3685	0.3725	**0.1290**	**0.1963**
average	0.5384	0.3363	0.3391	**0.2103**	**0.3359**

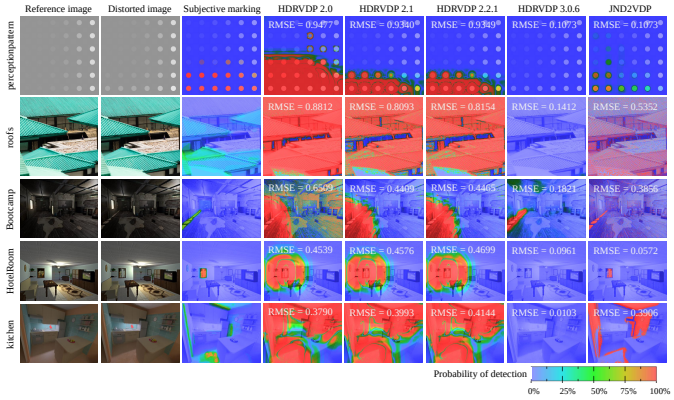


Fig. 16: Visual comparisons of the predicted visibility map results. JND2VDP represents the metric which is converted from our proposed JND model according to Eq. (19).

D. Comparison with VDP-type Metrics

As stated in Eqs. (1) and (2), the JND and VDP are two relevant concepts in that they both attempt to model the same underlying mechanism of the visual system - detection and discrimination. To enable a direct comparison between the JND model and VDP-type metrics, we need to convert the JND map into an approximate visibility map in the following manner:

$$O_{JND2VDP} \approx p\left((I_d - I_o)/F_{JND}(I_o)\right), \quad (19)$$

where $p(x)$ is the psychometric function. According to Eq. (17) in [48], $p(x)$ can be expressed as follows:

$$p(x) = 1 - \exp(\log(0.5)x^{\beta}), \quad (20)$$

where β is the slope of the psychometric function and the value set to $\beta = 3.5$ [49].

TABLE VII: Running Time Comparison.

Models	Yang2005 [21]	Zhang2005 [47]	Wu2013 [23]	Wu2017 [24]	Jakhetiya2018 [25]	Chen2020 [26]	Shen2021 [27]	Proposed
Time (s)	≈0.1372	≈5.8350	≈9.1736	≈0.9706	≈57.5209	≈0.6843	≈9.8281	≈0.2941

The JND-converted visibility map is denoted by JND2VDP in the following. In the experiments, we adopt the distortion visibility dataset provided in [50] as the benchmark. The JND2VDP is compared with HDR-VDP 2.0 [48], HDR-VDP 2.1 [48], HDR-VDP 2.2.1 [46, 48], and HDR-VDP 3.0.6 [48]. The details of comparison are illustrated as follows. For each scene in the dataset, there is a reference image, a distorted image, and a subjective marking map. The subjective marking map reflects the area and intensity of visible distortions in the distorted image. A larger intensity value in the subjective marking map means a larger probability of detecting the distortion in the distorted image. The values in the subjective marking map are normalized in the range [0,1] for computation. We take the subjective marking map as the ground-truth, and we compute the RMSE value between subjective marking map and the probability map produced by JND2VDP or other compared VDP metrics. A smaller RMSE value means a better prediction accuracy. The comparison results are shown in Table VI. From this table, we can observe that JND2VDP is only worse than the latest HDR-VDP 3.0.6 while slightly better than HDRVDP 2.0 [48], HDRVDP 2.1 [48], and HDRVDP 2.2.1 [46, 48]. In Fig. 16, we further visualize some results of the predicted visibility maps for reference. These results demonstrate that our predicted JND map has a good capability in distortion detection and discrimination.

E. Running Time

Besides the high prediction accuracy, an excellent JND model should also be computationally efficient. We test the running time of different JND models with the same setting and platform. The experiments are all conducted on a PC with an AMD Ryzen 7 4800H @ 2.9 GHZ and 16GB RAM. The software platform is MATLAB R2019a. The running times of different JND models are compared in Table VII. We present the running times in the unit of second (s) and a smaller value means a faster running speed. Our proposed JND model ranks the second place among all competitors with a running speed within 0.3ms for processing a 1200×800 image. Although Yang2005 is more efficient, the inferior prediction accuracy makes it unsuitable for using in practical applications.

F. Limitation

Although the proposed JND model can achieve better performance than the existing ones, it also has limitations. Our model is also based upon data obtained in one subjective experiment and its parameters are fully dependent on the experimental conditions of the subjective test. Thus, our JND model is only suitable for that viewing condition of test and cannot flexibility adapt for different viewing conditions. According to [34, 35], it is known that viewing conditions will have direct influence on the visually lossless threshold of images. The user studies in [34, 35] also show that higher peak luminance and shorter viewing distance will help users to more easily discover distortions. However, our work does not take into account the influence of viewing condition, i.e., we only focus on predicting the JND of images under a specific viewing condition setting. In other words, our work only accounts for image content while ignoring the influence of viewing condition. For example, the standard viewing conditions are mostly at 60 ppd while we set ppd=30 as an alternative to setup our subjective experiments. We admit that such an ignorance of viewing condition is an obvious limitation of our method. In the future, we will devote to considering viewing condition as an influential factor toward building a more flexible and effective JND prediction models in practical applications.

Another limitation is that the concept of a JND map does not account for the notion of energy or spatial pooling. For example, if we change a single pixel by the value from the JND map, the change is unlikely to be detected. But, if we change a large number of pixels (according to the JND map), the change is likely to be well-visible. The type of change will also have a strong effect on the visibility of changes. If we add the same value to all pixels (e.g., the minimum from the JND map), the change is unlikely to be detected because we are insensitive to low-frequency or direct component (DC) brightness changes. But, if we add salt-and-paper noise of the same amplitude across the image, the change is likely to be observed. The JND maps cannot distinguish between those cases.

IV. Conclusion

This paper has presented a novel top-down JND estimation model of natural images. It dedicates to estimating a CPL image first by exploiting the distribution characteristic of the cumulative normalized KLT coefficient energy and then calculating the difference map bettween the original image and CPL image as the JND map. The difference map well reflects the redundant micro-structure information which typically cannot be perceived by the HVS. Using such a difference map as the final JND map, the visual redundancies in the image can be better exploited. We have evaluated the performance of the proposed JND model explicitly with direct JND prediction and implicitly with two applications including JND-guided noise injection and JND-guided image compression. Experimental results have demonstrated that our proposed JND model can achieve better performance than several latest JND models. In addition, we also compare the proposed JND model with existing VDP metrics in terms of the capability in distortion detection and discrimination. The results indicate that our proposed JND model also has a good performance in this task.

REFERENCES

[1] W. Lin and G. Ghinea, "Progress and opportunities in modelling just-noticeable difference (JND) for multimedia," *IEEE Transactions on Multimedia*, 2021.

[2] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2025–2043, 2013.

[3] X. Zhang, S. Wang, K. Gu, W. Lin, S. Ma, and W. Gao, "Just-noticeable difference-based perceptual optimization for jpeg compression," *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 96–100, 2017.

[4] C.-M. Mak and K. N. Ngan, "Enhancing compression rate by just-noticeable distortion model for H.264/AVC," in *2009 IEEE International Symposium on Circuits and Systems*, 2009, pp. 609–612.

[5] X. Yang, W. Ling, Z. Lu, E. Ong, and S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Processing: Image Communication*, vol. 20, no. 7, pp. 662–680, 2005.

[6] S.-W. Jung, J.-Y. Jeong, and S.-J. Ko, "Sharpness enhancement of stereo images using binocular just-noticeable difference," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1191–1199, 2012.

[7] W. Lin, Y. Gai, and A. Kassim, "Perceptual impact of edge sharpness in images," *in IEE Proceedings of Vision, Image and Signal Processing*, vol. 152, pp. 215–223, 05 2006.

[8] C.-H. Chou and K.-C. Liu, "A perceptually tuned watermarking scheme for color images," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2966–2982, 2010.

[9] Y. Niu, J. Liu, S. Krishnan, and Q. Zhang, "Combined just noticeable difference model guided image watermarking," in *2010 IEEE International Conference on Multimedia and Expo*, 2010, pp. 1679–1684.

[10] Y. Niu, M. Kyan, L. Ma, A. Beghdadi, and S. Krishnan, "Visual saliency's modulatory effect on just noticeable distortion profile and its application in image watermarking," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 917–928, 2013.

[11] M. Bouchakour, G. Jeannic, and F. Autrusseau, "JND mask adaptation for wavelet domain watermarking," in *2008 IEEE International Conference on Multimedia and Expo*, 2008, pp. 201–204.

[12] Q. Cheng and T. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 273–284, 2001.

[13] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 9, pp. 2546–2559, 2016.

[14] X. Chen, G. Gao, D. Liu, and Z. Xia, "Steganalysis of LSB matching using characteristic function moment of pixel differences," *China Communications*, vol. 13, no. 7, pp. 66–73, 2016.

[15] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[16] S. A. Fezza, M.-C. Larabi, and K. M. Faraoun, "Stereoscopic image quality metric based on local entropy and binocular just noticeable difference," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 2002–2006.

[17] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranza, "Adaptive watermarking and tree structure based image quality estimation," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 311–325, 2014.

[18] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, 2013.

[19] S. Seo, S. Ki, and M. Kim, "A novel just-noticeable-difference-based saliency-channel attention residual network for full-reference image quality predictions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2602–2616, 2021.

[20] C.-H. Chou, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," in *Proceedings of 1994 IEEE International Symposium on Information Theory*, 1994, pp. 420–.

[21] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue pre-processing in video coding based on just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 6, pp. 742–752, 2005.

[22] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1648–1652, 2010.

[23] J. Wu, G. Shi, W. Lin, A. Liu, and F. Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1705–1710, 2013.

[24] J. Wu, L. Li, W. Dong, G. Shi, W. Lin, and C.-C. J. Kuo, "Enhanced just noticeable difference model for images with pattern complexity," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2682–2693, 2017.

[25] V. Jakhetiya, W. Lin, S. Jaiswal, K. Gu, and S. C. Guntuku, "Just noticeable difference for natural images using RMS contrast and feed-back mechanism," *Neurocomputing*, vol. 275, pp. 366–376, 2018.

[26] Z. Chen and W. Wu, "Asymmetric foveated just-noticeable-difference model for images with visual field inhomogeneities," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4064–4074, 2020.

[27] X. Shen, Z. Ni, W. Yang, X. Zhang, S. Wang, and S. Kwong, "Just noticeable distortion profile inference: A patch-level structural visibility learning approach," *IEEE Transactions on Image Processing*, vol. 30, pp. 26–38, 2021.

[28] H. Wang, L. Yu, J. Liang, H. Yin, T. Li, and S. Wang, "Hierarchical predictive coding-based JND estimation for image compression," *IEEE Transactions on Image Processing*, vol. 30, pp. 487–500, 2021.

[29] J. Wu, G. Shi, and W. Lin, "A survey of visual just noticeable difference estimation," *Front.Comput.Sci.*, 2019.

[30] Y. Jia, W. Lin, and A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 820–829, 2006.

[31] Z. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 337–346, 2009.

[32] S.-H. Bae and M. Kim, "A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3227–3240, 2014.

[33] W. Wan, J. Wu, X. Xie, and G. Shi, "A novel just noticeable difference model via orientation regularity in DCT domain," *IEEE Access*, vol. 5, pp. 22 953–22 964, 2017.

[34] A. Mikhailiuk, N. Ye, and R. K. Mantiuk, "The effect of display brightness and viewing distance: a dataset for visually lossless image compression," *Electronic Imaging*, 2021.

[35] N. Ye, K. Wolski, and R. K. Mantiuk, "Predicting visible image differences under varying display brightness and viewing distance," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5429–5437.

[36] L. Jin, J. Y. Lin, S. Hu, H. Wang, P. Wang, I. Katsavounidis, A. Aaron, and C.-C. J. Kuo, "Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis," *Electronic Imaging*, vol. 2016, no. 13, pp. 1–9, 2016.

[37] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-

VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 1–14, 2011.
[38] X. Zhang, C. Yang, X. Li, S. Liu, H. Yang, I. Katsavounidis, S.-M. Lei, and C.-C. J. Kuo, "Image coding with data-driven transforms: Methodology, performance and potential," *IEEE Transactions on Image Processing*, vol. 29, pp. 9292–9304, 2020.
[39] C. Yang, X. Zhang, P. An, L. Shen, and C.-C. J. Kuo, "Blind image quality assessment based on multi-scale KLT," *IEEE Transactions on Multimedia*, vol. 23, pp. 1557–1566, 2021.
[40] X. Zhang, S. Kwong, and C.-C. J. Kuo, "Data-driven transform based compressed image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
[41] R. Timofte et al., "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1110–1121.
[42] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1122–1131.
[43] Z. Yang, L. Tian, and C. Li, "A fast video shot boundary detection employing OTSU's method and dual pauta criterion," in *2017 IEEE International Symposium on Multimedia (ISM)*, 2017, pp. 583–586.
[44] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, pp. 591–611, 1965.
[45] M. Alfawzan, "Methods for estimating the parameters of the weibull distribution," *Fawzan*, no. 10, 2000.
[46] M. Narwaria, R. K. Mantiuk, M. Silva, and P. L. Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, p. 010501, 2014.
[47] X. Zhang, W. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortion," *Signal Processing*, vol. 85, no. 4, pp. 795–808, 2005.
[48] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 1–14, 2011.
[49] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Electronic Imaging*, 1992.
[50] K. Wolski, D. Giunchi, N. Ye, P. Didyk, K. Myszkowski, R. Mantiuk, H.-P. Seidel, A. Steed, and R. K. Mantiuk, "Dataset and metrics for predicting local visible differences," *ACM Transactions on Graphics*, vol. 37, pp. 1–14, 2018.

Qiuping Jiang (M'18) is currently an Associate Professor with Ningbo University, Ningbo, China. He received the Ph.D. degree in Signal and Information Processing from Ningbo University in 2018. From Jan. 2017 to May 2018, he was a visiting student with Nanyang Technological University, Singapore. His research interests include image processing, visual perception, and computer vision. He received the Distinguished Youth Scholar Funding of Zhejiang Natural Science Foundation, the Best Paper Honorable Mention Award of the *Journal of Visual Communication and Image Representation*, and the Excellent Doctoral Dissertation Award of Zhejiang Province. He also serves as the Associate Editor of *Journal of Electronic Imaging* and *APSIPA Trans. on Information and Signal Processing*, and the Area Chair/Session Chair/PC member for IJCAI/AAAI/ACM-MM/ICME/ICIP/APSIPA-ASC.

Zhentao Liu is currently an forth-year undergraduate student major in Communication Engineering with the School of Information Science and Engineering, Ningbo University, China. His research interests include image processing, image quality assessment, and visual perception modeling.

Shiqi Wang received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in Computer Application Technology from the Peking University under the supervision of Prof. Wen Gao, in 2014. From Mar. 2014 to Mar. 2016, He was a Postdoc Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. From Apr. 2016 to Apr. 2017, He was with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. His research interests include video compression, image/video quality assessment, and image/video search and analysis.

Feng Shao received the B.S. and Ph.D. degrees in Electronic Science and Technology from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. In 2012, he was a visiting scholar with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, image quality assessment, and immersive media computing.

Weisi Lin (F'16) is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He received the Bachelor's degree in Electronics and then a Master's degree in Digital Signal Processing from Sun Yat-Sen University, Guangzhou, China, and the Ph.D. degree in Computer Vision from King's College, London University, UK. His research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.

He is a Fellow of the IEEE and IET, an Honorary Fellow of the Singapore Institute of Engineering Technologists, and a Chartered Engineer in U.K. He was the Chair of the IEEE MMTC Special Interest Group on Quality of Experience. He was awarded as the Distinguished Lecturer for IEEE Circuits and Systems Society in 2016-2017. He served as a Lead Guest Editor for a Special Issue on Perceptual Signal Processing of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING in 2012. He also has served or serves as an Associate Editor for IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Multimedia, IEEE Signal Processing Letters, and Journal of Visual Communication and Image Representation.