Copyright ⓒ 2024 KSII

Adversarial Complementary Learning for Just Noticeable Difference Estimation

Dong Yu[1], Jian Jin[2*], Lili Meng[1*], Zhipeng Chen[3], and Huaxiang Zhang[1]
[1] School of Information Science and Engineering, Shandong Normal University
Jinan, 250014, China
[e-mail: yudong11222@163.com, mengll_83@hotmail.com, huaxzhang@hotmail.com]
[2] School of Computer Science and Engineering, Nanyang Technological University
Singapore, 639798, Singapore
[e-mail: jian.jin@ntu.edu.sg]
[3] Department of Computer Science, Tangshan Normal University
Tangshan, 063000, China
[e-mail: chzhpeng@hotmail.com]
*Corresponding author: Jian Jin, Lili Meng

Received August 10, 2023; revised November 1, 2023; revised December 24, 2023; accepted January 20, 2024; published February 29, 2024

Abstract

Recently, many unsupervised learning-based models have emerged for Just Noticeable Difference (JND) estimation, demonstrating remarkable improvements in accuracy. However, these models suffer from a significant drawback is that their heavy reliance on handcrafted priors for guidance. This restricts the information for estimating JND simply extracted from regions that are highly related to handcrafted priors, while information from the rest of the regions is disregarded, thus limiting the accuracy of JND estimation. To address such issue, on the one hand, we extract the information for estimating JND in an Adversarial Complementary Learning (ACoL) way and propose an ACoL-JND network to estimate the JND by comprehensively considering the handcrafted priors-related regions and non-related regions. On the other hand, to make the handcrafted priors richer, we take two additional priors that are highly related to JND modeling into account, i.e., Patterned Masking (PM) and Contrast Masking (CM). Experimental results demonstrate that our proposed model outperforms the existing JND models and achieves state-of-the-art performance in both subjective viewing tests and objective metrics assessments.

Keywords: Just Noticeable Difference (JND), convolutional neural networks, Human Visual System (HVS).

This work was supported in part by the NSF of Shandong Province under Grant ZR2020MF042 and Grant ZR2022MF346; Science and Technology Plan Project of Tangshan Science and Technology Bureau Tangshan Foundation Innovation Team of Digital Media Security under Grant 21130212D.

http://doi.org/10.3837/tiis.2024.02.009 **ISSN : 1976-7277**

1. Introduction

Just Noticeable Difference (JND) refers to the maximum image pixel change that the Human Visual System (HVS) [1] cannot perceive, which reflects the visual redundancy of the HVS. Therefore, JND models are commonly used for estimating the visual redundancy that existed in the image/video. Hereby, JND models are widely used in image/video processing fields, such as perceptual image/video compression [2] [3] [4] [5] [6], quality assessment [7] [8] [9], privacy preserving [10] [11] [12], watermarking [13] [14], and so on. JND modeling has been studied for decades. The main goal of JND modeling is to accurately estimate the visual redundancy of the HVS in terms of different visual content. Currently, the existing JND models can be approximately classified into two main categories, that is, HVS-inspired JND models [15] [16] [17] [18] [19] [20] [21] [22] and learning-based JND ones [23] [24] [25] [26] [27] [28] [29] [30].

HVS-inspired JND models are mainly mathematically formulated by developing various maskings, which are highly related to the characteristics of the HVS. For instance, contrast [16], concealment effect [21], luminance adaptation [15], pattern complexity [20], foveated effect [17], and visual attention [22] were formulated as various maskings for modeling JND. Such kinds of JND models commonly have good interpretability. However, as the perceptual mechanism of HVS is still not fully understood, this limits the development of such kinds of JND models.

With the great success achieved by deep learning, many efforts [23] [24] [25] [26] [27] [28] [29] [30] are made to model JND via learning-based methods, termed learning-based JND models. It significantly improved the accuracy of JND modeling. Especially, Wu et al. [27] and Jin et al. [23] built an unsupervised learning JND generative network and used image quality assessment (IQA) to guide training. Although their performances have been improved, the generative network they used heavily relies on handcrafted priors. This restricts the information for estimating JND simply extracted from regions that are highly related to handcrafted priors, while information from the rest of the regions is disregarded, limiting their accuracy in JND estimation. Specifically, the JND estimated by their models mainly focuses on the regions guided by the handcrafted priors, which causes the estimate of the JND in the non-related handcrafted priors regions to be inaccurate. Inspired by Zhang et al. [31], we introduce Adversarial Complementary Learning (ACoL) into the modeling of JND, termed ACoL-JND. It comprehensively considers the JND of handcrafted priors-related regions and non-related regions in the image. In addition, the pattern complexity (PC) [20] of images were used as the handcrafted priors in [27] and [23] to guide the generation of JND and achieved good performance. However, PC is only a representation of pattern masking and cannot reflect the spatial masking effect. Hence, we introduce a spatial masking, composed of pattern masking and contrast masking, into the JND modeling to enrich the handcrafted priors. The final results show that our model is highly consistent with HVS and has reached state-of-the-art performance.

1.1 Contributions

The following is a summary of the paper's main contributions:

- We propose a novel JND generation network model by adopting ACoL, termed as ACoL-JND. The ACoL-JND model is composed of two parallel convolutional networks, which extracts different information in images by using dynamic erasure design. This allows the information of the handcrafted priors-related regions and non-related regions is complementally utilized to estimate JND.

- To enrich the handcrafted priors, two additional priors that are highly related to JND modeling are considered, including pattern masking (PM) and contrast masking (CM). They form spatial masking and serve as prior knowledge to guide the network for JND generation.
- We compared our model with other models in detail through subjective observation testing and objective Image Quality Assessment (IQA) evaluation, demonstrating that our proposed model improves the accuracy of JND modeling.

1.2 Organization

The remainder of the paper will be described in the following sections. Section 2 introduces the related work of JND models. Section 3 provides an introduction to the specifics of ACoL-JND. In Section 4, we present a series of studies (i.e., detailed comparison, ablation study, subjective observation test, objective IQAs evaluation, complementary features analysis). Section 5 concludes the paper.

2. Related Works

In this section, we mainly review the existing JND models. In addition, as IQA-relevant techniques are involved in this work, we also review the IQA metrics.

2.1 HVS-Inspired JND Models

Bae et al. [16] proposed a JND model based on a novel measure of texture complexity by considering the visual patterns and contrast intensity. It revealed that more signal changes can be tolerated by the HVS at the disordered texture regions in the image. After that, Wu et al. [21] assumed that the concealment effect of disordered regions in the image is higher than that of ordered regions and built a model for JND estimation. Meanwhile, Bae et al. [15] combined the influence of DCT domain frequency and image background luminance on the luminance adaptation effect and proposed a DCT-based JND model. Inspired by cognitive science that visual content can be represented by the image's repetitive patterns extracted by HVS, Wu et al. [20] introduced PC into the JND modeling and improved the accuracy of JND estimation. Afterward, Chen et al. [17] observed that the JND increased when visual eccentricity became large, and they proposed foveated masking for the JND modeling. Additionally, Zeng et al. [22] also took visual saliency into account and used it to scale the JND. All these models can predict the JND for each pixel of the image while their accuracy is limited due to the incomprehensible of the HVS.

2.2 Learning-Based JND Models

Jin et al. [24] first proposed the picture-wise JND and developed a JND dataset (e.g., MCL-JCI) for learning the picture-wise JND. Subsequently, Liu et al. [25] regarded picture-wise JND as a classification problem and learned a picture-wise JND profile for picture compression. After that, Wang et al. [26] built a large-scale JND video dataset by taking video compression distortion into account, termed VideoSet. However, these two JND datasets only considered the distortion of image and video compression. To cover more distortions, Liu et al. [32] built the first comprehensive JND dataset with multiple distortion types. Meanwhile, to estimate the satisfied user ratio (SUR) and video-wise JND, Zhang et al. [28] introduced temporal and spatial details into the proposed JND and SUR model. However, all these learning-based JND models reviewed above were able to predict the picture/video-wise, while

they cannot estimate the JND for each pixel. To estimate the JND for each pixel, Wu et al. [27] utilized convolutional neural networks to generate the JND for each pixel, while the reasonability of the generated JND was measured with an IQA metric (e.g., SSIM). Besides, the handcrafted prior PC was used for guiding JND generation. All these designs estimated the JND in an unsupervised learning way. After that, to make the handcrafted prior richer, Jin et al. [23] introduced visual attention (VA) [33] in JND generation. Besides, the characteristics of full RGB channels were considered. To better assess the generated JND they also proposed an adaptive IQA (A-IQA) module for assessing the generated JND, which largely improved the accuracy of JND. The unsupervised learning-based JND models reviewed above significantly improve the JND modeling for each pixel. However, they heavily relied on handcrafted priors, resulting in extracted information solely from regions highly related to handcrafted priors, while ignoring information from the rest of the regions. Considering that JND is a result determined by a whole perception of the image, all the relevant information should be considered when modeling without discrimination.

2.3 IQA

Image quality assessment (IQA) metrics are developed to replace humans in accurately evaluating the quality of images. After decades of studies, the accuracy of the IQA metrics has been largely improved. There are several typical IQA metrics, that are widely used in the image processing field, including Structural Similarity Index Measure (SSIM) [34], Gradient Magnitude Similarity Deviation (GMSD) [35], Normalized Laplacian Pyramid Distance (NLPD) [36], Multi-Scale Structural Similarity Index Measure (MS-SSIM) [37], Feature Similarity Index Measure (FSIM) [38], Most Apparent Distortion (MAD) [39], Visual Information Fidelity (VIF) [40], Deep Image Structure and Texture Similarity (DISTS) [41], Visual Saliency Induced (VSI) [42], Learned Perceptual Image Patch Similarity (LPIPS) [43], Complex Wavelet SSIM (CW-SSIM) [44], Gradient Similarity (GSM) [45], and so on.

However, these IQA metrics can only perform well for specific types of distortion. There are at least 17 types of distortion (e.g., contrast change, mean shift, spatially correlated noise, and so on) as studied in the TID2008 [46] dataset. Besides, with the continuous development of multimedia technology, more and more new types of distortion have been generated. At present, there is no single IQA that can perform well on all types of distortion. To solve this problem, Liu et al. [47] proposed a multi-method fusion (MMF) algorithm. They divided all distortion types in the TID2008 [46] dataset into five groups. Then, they combined multiple IQAs based on machine learning techniques to achieve the best performance. After that, Jin et al. [23] assigned the best combination consisting of three IQAs to each distortion group by training a distortion-type classifier using ResNets [48] networks and they proposed an adaptive image quality assessment (A-IQA) module to replace humans in assessing the reasonability of the generated JND and achieved good performance. Inspired by this, this paper will use the A-IQA module to guide network training to improve the accuracy of JND estimation.

3. Proposed ACoL-JND Model

3.1 Architecture

The existing unsupervised learning-based JND models [27] [23] have a heavy reliance on handcrafted priors for guidance. However, this restricts the information for estimating JND simply extracted from regions that are highly related to handcrafted priors, while information from the rest of the regions is disregarded. This has caused the limitation of their ability to

estimate JND. Zhang et al. [31] proposed a novel adversarial complementary learning method for weakly supervised object localization and achieved great success. Inspired by this, we adopt adversarial complementary learning to fully extract the complementary information of the handcrafted priors-related regions and non-related regions to improve the performance of Jin et al.'s [23] JND model. Meanwhile, we modify the network of [31] to make it more suitable for JND estimation, and more details are exhibited in **Fig. 1**. Our model is composed of four convolutional networks (i.e., CNN1, CNN2, CNN3, and CNN4) and two upper sampling layers (i.e., Up1, Up2). Besides, four operations (i.e., Stack, Erasing, Element-Wise addition, and Thresholding) and an adaptive IQA (A-IQA) [23] module are included. Among them, CNN1, CNN2, CNN3, and CNN4 are used to extract image features. Up1 and Up2 are used to upsample the features to meet the size requirements of the next operation. A leaky rectified linear unit (LReLU) [49] rectifier is used to activate each convolution layer. The specific details of the network structure are shown in **Table 1**. Then, the stack operation Ⓢ is used to stack two tensors according to a certain dimension. The erasing operation ⊝ is used to dynamically erase image characteristics according to a mask. The element-wise addition operation ⊕ is used to inject the generated JND map into the original image. The thresholding operation is used to generate a mask according to the specified threshold. Finally, the A-IQA [23] module evaluates the quality of distorted images and optimizes the network training as a loss. The specific architecture of the ACoL-JND model is shown in **Fig. 1**.

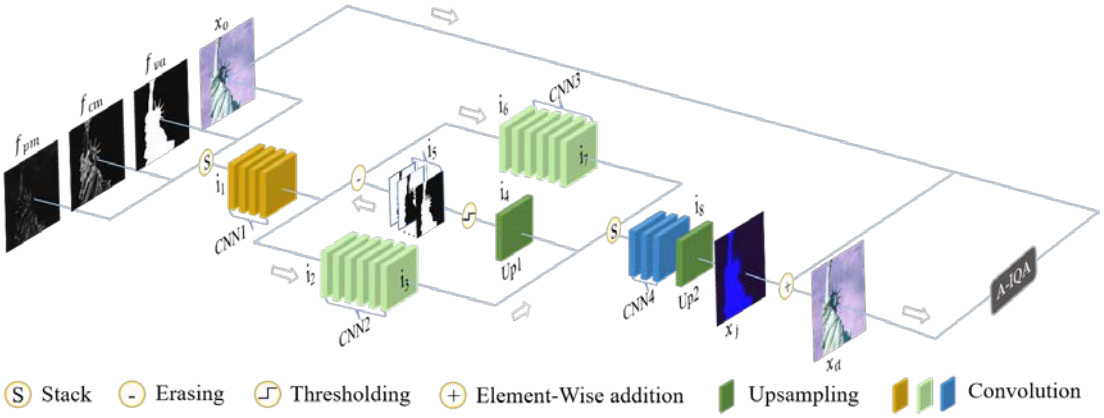


Fig. 1. The architecture of the ACoL-JND network.

First, we use f_{pm} [20], f_{cm} [20], and f_{va} [33] (i.e., we retain the f_{va} of Jin et al.'s [23] JND model) as the prior knowledge, which is stacked with the original image x_o as the input i_1 of CNN1. As shown in the following formula.

$$\{x_o, f_{pm}, f_{cm}, f_{va}\} = \{x_o\} Ⓢ \{f_{pm}, f_{cm}, f_{va}\} \tag{1}$$

After passing through CNN1, features i_2 is obtained. i_2 is then fed into CNN2 to obtain the features i_3. i_3 is upsampled by Up1 to obtain the features i_4. i_4 is thresholded to get the masking i_5. Later, the masking i_5 is used to erase the features extracted by CNN2 in i_2 to obtain the features i_6. Therefore, we have the following formula.

$$i_6 = i_2 ⊝ i_5 \tag{2}$$

After that, the features i_6 is fed into CNN3 to obtain the features i_7. i_3 and i_7 are stacked and fed into CNN4 to obtain the features i_8. i_8 is upsampled to generate the JND map x_j. Subsequently, the JND map x_j is injected into the original image x_o. Then, we get a distorted image x_d and this process is formulated as.

$$x_d = x_o \oplus x_j \tag{3}$$

Eventually, we use the A-IQA [23] module to evaluate the quality of the distorted image x_d with the original image x_o as a reference and use it as a loss to optimize the network training.

Table 1. The detailed architecture of the main components in the ACoL-JND network

Network	Name	Input Shape	Operation	Output Shape	Activation
CNN1	1 E	32, 6, 176, 176	Conv. (3 × 3)	32, 8, 176, 176	LReLU
	2 E	32, 8, 176, 176	Conv. (3 × 3)	32, 16, 176, 176	LReLU
	3 E	32, 16, 176, 176	Conv. (3 × 3)	32, 16, 176, 176	LReLU
	4 E	32, 16, 176, 176	Conv. (3 × 3)	32, 32, 176, 176	LReLU
CNN2 / CNN3	1 E	32, 32, 176, 176	Conv. (3 × 3)	32, 32, 176, 176	LReLU
	2 E	32, 32, 176, 176	Conv. (3 × 3)	32, 64, 88, 88	LReLU
	3 E	32, 64, 88, 88	Conv. (3 × 3)	32, 128, 88, 88	LReLU
	4 E	32, 128, 88, 88	Conv. (3 × 3)	32, 64, 44, 44	LReLU
	5 E	32, 64, 44, 44	Conv. (3 × 3)	32, 32, 44, 44	LReLU
	6 E	32, 32, 44, 44	Conv. (3 × 3)	32, 32, 44, 44	LReLU
CNN4	1 E	32, 64, 44, 44	Conv. (3 × 3)	32, 32, 44, 44	LReLU
	2 E	32, 32, 44, 44	Conv. (3 × 3)	32, 16, 44, 44	LReLU
	3 E	32, 16, 44, 44	Conv. (3 × 3)	32, 3, 44, 44	LReLU
Up1	—	32, 32, 44, 44	Bilinear	32, 32, 176, 176	—
Up2	—	32, 3, 44, 44	Bilinear	32, 3, 176, 176	—

3.2 Spatial Masking Based JND Estimation

The existing JND models [27] [23] use the handcrafted prior i.e., PC [20] to guide the network for JND estimation. However, they cannot accurately estimate the JND, resulting in obvious noise in the object edges and background regions, due to PC only being a representation of pattern masking and not reflecting the spatial masking effect. Hence, spatial masking highly related to JND modeling is considered, which is composed of pattern masking f_{pm} and contrast masking f_{cm} from [20]. It can make the handcrafted priors richer and guide the network to generate JND. Meanwhile, pattern masking f_{pm} and contrast masking f_{cm} play different roles in the process of JND estimation. Among them, pattern masking f_{pm} mainly guides the network to estimate JND in irregular regions. And contrast masking f_{cm} has a better performance in guiding the network to estimate JND in object regular edge regions. Generally, pattern masking f_{pm} and contrast masking f_{cm} exist simultaneously in one image. Wu et al. [20] believe that one of the two masking effects plays a leading role, and regard the dominant masking effect as the final spatial masking. This way will ignore the role of another masking effect in spatial masking. In this paper, we consider both masking effects in our JND modeling and retain [23]'s visual attention f_{va} [33] to improve the accuracy of JND estimation.

3.3 Loss Function

ACoL-JND model mostly inherits Jin et al.'s [23] loss function. As mentioned above, we use the A-IQA [23] module to evaluate the quality of distorted images and optimize network training as a loss. Thus, the formula is as follows.

$$L_1 = \mathrm{Q}(x_o, x_d) \tag{4}$$

Then, to guarantee that the generated JND cannot be perceived by HVS, we prefer to control the JND injection to the high gradient region of the image. Consequently, we inherit [23] as follows.

$$L_2 = ln(M^2 + N^2 + h_0) - ln\,(2 \bullet M \bullet N + h_0) \tag{5}$$

The gradient of the original image is represented by M, and the value of the JND map is represented by N. In addition, to prevent the denominator from being zero, we add a constant h_0, $h_0 = 0.001$.

Ultimately, we have a total loss function as follows.

$$L = \delta \bullet L_1 + \eta \bullet L_2 \tag{6}$$

L is the total loss of the network. Refer to [27], [23], set $\delta = 1$ and $\eta = 0.1$.

3.4 Implementation Details

In this paper, we evaluate our performance against the anchor methods in COCO2017 [50] and CSIQ [51] datasets. Among them, randomly select 1000 pictures from COCO2017 [50] to train the model. The threshold value of each channel is the mean value of each channel after Up1 sampling. The batch size is 32, and the learning rate is 10^{-5}. Adam [52] is used to optimize network training. Four models (i.e., Wu2017 [20], Wu2020 [27], Jiang2022 [53], Jin2022 [23]) are used as anchors to compare with our proposed model. We inject the JND map into the original image as noise according to the following formula.

$$D = O \oplus (\lambda \bullet R \bullet J) \tag{7}$$

R is a random matrix, whose value only includes positive one and negative one. JND maps generated by different JND models are represented by J. Then, the magnitude of the JND map J is adjusted by λ to inject the original image O to obtain a distorted image D.

4. Experiments

In this section, we compare our proposed model with four anchor methods. Then, we conduct ablation studies to further illustrate the reasonability of our proposed model. The PSNR of distorted images in **Fig. 2** (a2) - (b4) is adjusted to 26.06 dB via (7). In addition, subjective observation tests and objective IQA evaluations are conducted to compare the performance of the proposed model with the anchor models. Finally, we analyze the crucial complementary features generated in the network.

4.1 JND Model Comparison

As shown in **Fig. 2**, four anchor methods (a2) - (a5) (i.e., Wu2017 [20], Wu2020 [27], Jiang2022 [53], Jin2022 [23]) are compared with our proposed model (a6). At first, Wu2017 [20] further refine the order and disorder regions of the image using pattern complexity. Thus, the high-complexity regions are injected with more noise resulting in obvious distortion, such as the face of the Statue of Liberty. Then, Wu2020 [27] takes pattern complexity as the handcrafted priors to guide the network for JND estimation. However, we can find that they overestimate JND in the sky background, resulting in obvious noise in such region. Meanwhile, Jiang2022 [53] takes the difference between the original image and the Critical Perceptual Lossless (CPL) image directly generated from a top-down perspective as the predicted JND. However, they overestimate the JND at the edge of the object. The obvious noise is found in the outline of the Statue of Liberty. Finally, although Jin2022 [23] achieved good results through RGB full channel modeling, the noise of the Statue of Liberty's head is still more obvious than the model we proposed. In summary, our model has great performance in

estimating the JND of the smooth background and the object edge regions.

Fig. 2. Detailed comparison among different models.

4.2 Ablation Study

We design ablation studies to illustrate our proposed model. As shown in **Fig. 2**, (b1) is the BL, (b2) is the BL+ACoL, (b3) is the BL+PM+CM**,** and (b4) is the BL+ACoL+PM+CM. We regard the JND model of Jin2022 [23] as BL, and BL+ACoL represents the BL used ACoL network. Then, BL+PM+CM represents using PM and CM to replace PC. Finally, the model proposed in this paper integrates ACoL and PM+CM, which are recorded as BL+ACoL+PM+CM, i.e., Ours. It can be seen from **Fig. 2** that the noise on the face of the Statue of Liberty in BL is more obvious than in others. Although BL+ACoL has less noise on the face of the Statue of Liberty than BL, the background noise has not decreased significantly.

The noise of BL+PM+CM in the background is reduced, but the noise of the Statue of Liberty′s face is obvious. Finally, we integrated ACoL and PM+CM to get BL+ACoL+PM+CM. It takes

into account the noise in the face region of the Statue of Liberty and the background region, so it achieves the best performance.

4.3 Subjective Observation Test

To further compare the performance of anchors (i.e., Wu2017 [20], Wu2020 [27], Jiang2022 [53], Jin2022 [23]) with our proposed model, we conduct a subjective observation test experiment. P1 to P12 in **Fig. 3** is a thumbnail of the test image from the CSIQ [51] dataset. Then, 21 testers were invited to participate in the subjective test. During the experiment, the distorted images of anchors and our model appear randomly on both sides of the screen. The correspondence between images and models is unknown to testers. The experimental environment's configurations refer to ITU-R BT.500-11 [54]. Testers scored according to the subjective quality of images, and the scoring criteria are shown in **Table 2**.

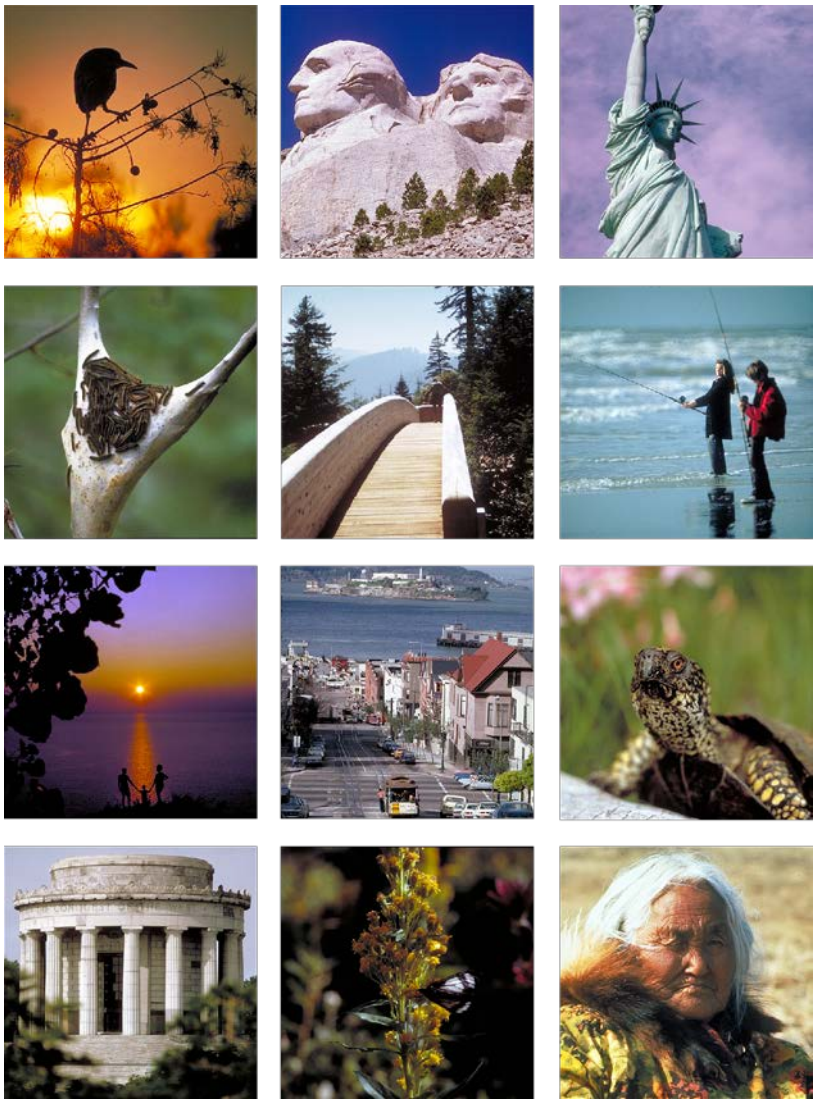


Fig. 3. Thumbnails of subjective observation tests.

Table 2. Scoring rules of image quality for subjective observation test

Description	A lot better	Better	Slightly better	Alike	Slightly better	Better	A lot better
Score	3	2	1	0	-1	-2	-3
Side	Left			Center	Right		

The results of the subjective observation test are shown in **Table 3**. The average value and standard deviation are denoted by 'Mean' and 'Std' respectively. A positive 'Mean' value indicates that the JND model on the upper side is superior to the JND model on the lower side. And the degree of goodness increases as the 'Mean' value increases. The consistency of the testers' judgments is reflected by the 'Std' value. Meanwhile, the higher the consistency, the closer the 'Std' is to zero. **Table 3** includes the comparison of our model with four anchors and the comparison of four models (i.e., BL (i.e., Jin2022 [23]), BL+ACoL, BL+PM+CM, Ours (i.e., BL+ACoL+PM+CM)) in the ablation study. We can find that most of the 'Mean' values in the table are positive. In addition, the average values of 'Mean' in the last row in **Table 3** are all positive. These findings demonstrate that our proposed model performs better than others in estimating JND.

Table 3. Subjective observation test results

Index	**Ours VS. Wu2017** [20]		**Ours VS. Wu2020** [27]		**Ours VS. Jiang2022** [53]		**Ours VS. Jin2022** [23]	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
P1	0.43	1.05	1.14	1.17	0.90	1.02	0.52	0.91
P2	0.67	0.99	0.71	1.20	0.67	1.04	0.05	0.65
P3	2.05	0.79	1.33	1.04	2.05	1.13	0.57	0.90
P4	1.33	1.08	1.14	0.77	1.76	0.97	0.52	1.33
P5	0.90	0.92	1.10	1.27	1.10	1.27	0.52	0.96
P6	0.86	0.94	0.76	1.34	1.90	0.92	0.33	1.04
P7	1.38	0.90	1.14	1.04	0.33	1.49	1.19	1.01
P8	0.86	1.39	1.10	1.11	1.52	1.14	0.71	1.20
P9	0.81	1.14	1.24	1.02	1.43	0.95	0.14	0.77
P10	1.14	1.25	1.43	0.90	1.10	0.97	0.57	0.73
P11	1.10	1.23	0.76	1.44	0.95	1.17	1.10	1.02
P12	0.90	1.11	0.86	0.89	1.10	1.27	0.71	0.70
Average	**1.04**	—	**1.06**	—	**1.23**	—	**0.58**	—
Index	**BL+ACoL VS. BL**		**BL+PM+CM VS. BL**		**Ours VS. BL+ACoL**		**Ours VS. BL+PM+CM**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
P1	0.10	0.81	0.19	0.79	0.57	0.95	0.57	1.05
P2	0.24	0.92	0.05	0.49	0.19	0.96	0.05	0.72
P3	0.52	0.96	0.19	0.66	0.90	0.75	0.33	0.64
P4	-0.19	1.14	0.10	0.75	0.57	0.79	0.14	0.99
P5	0.62	1.05	0.67	0.94	0.29	0.63	0.14	0.71
P6	0.57	0.49	0.48	1.05	0.10	0.61	0.10	0.68
P7	0.29	0.82	-0.10	1.02	0.71	0.98	0.76	0.97

P8	0.48	0.73	0.52	0.85	0.14	0.47	0.14	0.77
P9	0.05	0.72	0.38	0.79	0.19	0.59	-0.05	0.79
P10	0.05	0.72	0.33	0.89	0.10	0.68	0.29	0.55
P11	0.19	0.73	0.90	0.75	0.00	0.62	0.14	0.94
P12	0.57	0.66	0.57	0.90	0.10	0.61	0.24	0.81
Average	**0.29**	—	**0.36**	—	**0.32**	—	**0.24**	—

4.4 Objective IQAs Evaluation

In addition to subjective observation testing, we also utilize IQAs to objectively evaluate our proposed ACoL-JND model compared to four anchors. Meanwhile, the BL+ACoL and BL+PM+CM models used in the ablation study are also compared with our proposed model. Then, three advanced IQA evaluation metrics are used for objective estimation testing, including Normalized Laplacian Pyramid Distance (NLPD) [36], Most Apparent Distortion (MAD) [39], and Visual Information Fidelity (VIF) [40].

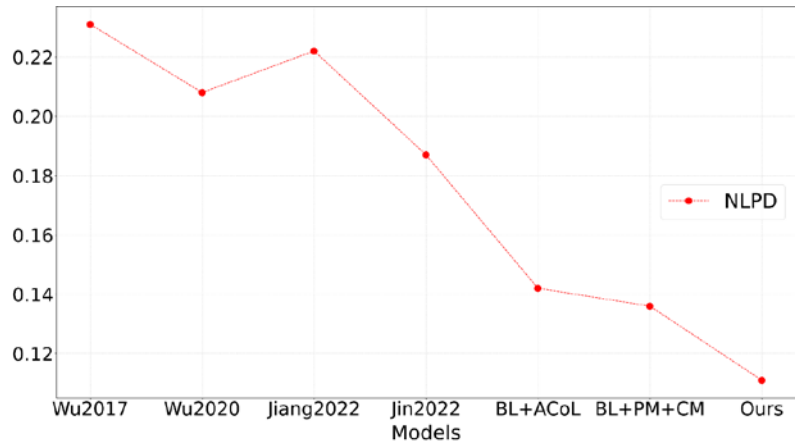


Fig. 4. Comparison of NLPD values of different models.

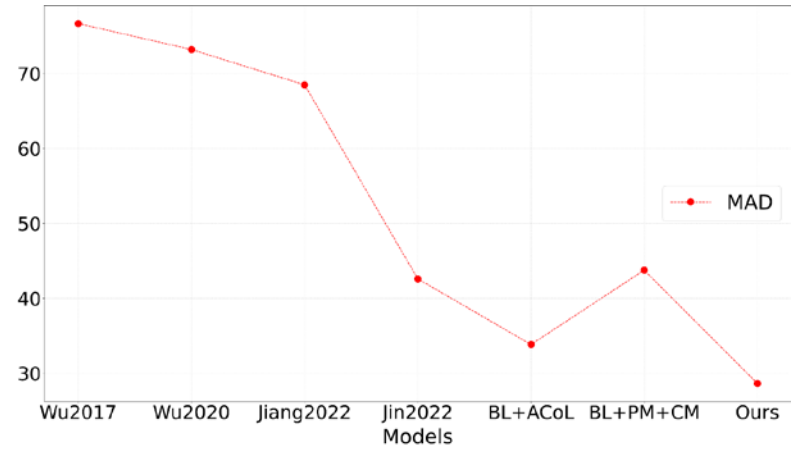


Fig. 5. Comparison of MAD values of different models.

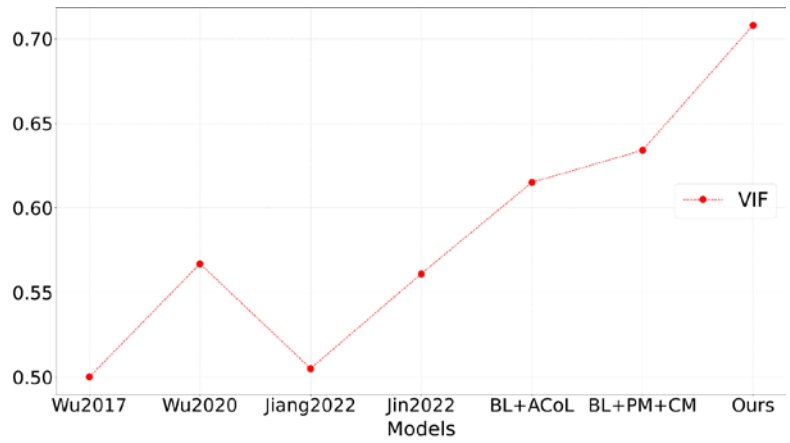


Fig. 6. Comparison of VIF values of different models.

We utilize these three objective IQA evaluation metrics to test our proposed model against other models on all images in the CSIQ [51] dataset. The average results are shown in **Fig. 4** to **Fig. 6**. At first, **Fig. 4** shows the results of NLPD, the smaller the value the better the performance of the JND model. From the results, we can find that the NLPD value of the ACoL-JND model is the lowest among all models. It indicates that the performance of our proposed model exceeds that of other models. Meanwhile, the MAD results shown in **Fig. 5** also indicate that smaller values represent better model performance. Our proposed ACoL-JND model also achieved the best performance compared to other models. Finally, **Fig. 6** shows the results of VIF. Unlike the NLPD and MAD mentioned above, the higher VIF value the better the performance of the model. From this result, it can be seen that the VIF value of our proposed model is higher than that of other models. This result is consistent with the results of NLPD and MAD, illustrating that our proposed ACoL-JND model has surpassed other models to achieve state-of-the-art performance.

We compare the proposed model with other models using objective IQAs evaluation and subjective observation test mentioned above. The results demonstrate that our model has higher accuracy in estimating image JND than other models.

4.5 Complementary Features Analysis

To further demonstrate the proposed ACoL-JND model, we display the crucial features generated during the **Fig. 1** network estimation of JND. These features are shown in **Fig. 7**. **Fig. 7** (a1) - (b1) are from the features i_2 of **Fig. 1**. The **Fig. 7** (a2) - (b2) are from the features i_4 of **Fig. 1**. The **Fig. 7** (a3) - (b3) are from the masking i_5 of **Fig. 1**. The **Fig. 7** (a4) - (b4) are from the features i_6 of **Fig. 1**. Except for (a3) and (b3), **Fig. 7** shows the results after normalization.

At first, we use CNN1 in **Fig. 1** to extract the initial features and obtain the i_2, which are shown in **Fig. 7** (a1) and (b1). Then, use CNN2 in **Fig. 1** to further extract the features and upsample by Up1 to obtain the i_4 as shown in **Fig. 7** (a2) and (b2). We can find that these features focus on regions that are highly related to handcrafted priors while ignoring non-related handcrafted priors regions. Subsequently, the i_4 is thresholded to generate the masking i_5 as shown in **Fig. 7** (a3) and (b3). The i_5 is used to erase some information extracted by CNN2 in the i_2 to obtain the i_6 as shown in **Fig. 7** (a4) and (b4). The CNN3 extracts the information of the non-related handcrafted priors regions (a4) and (b4) of the i_2 as complementary information of i_3.

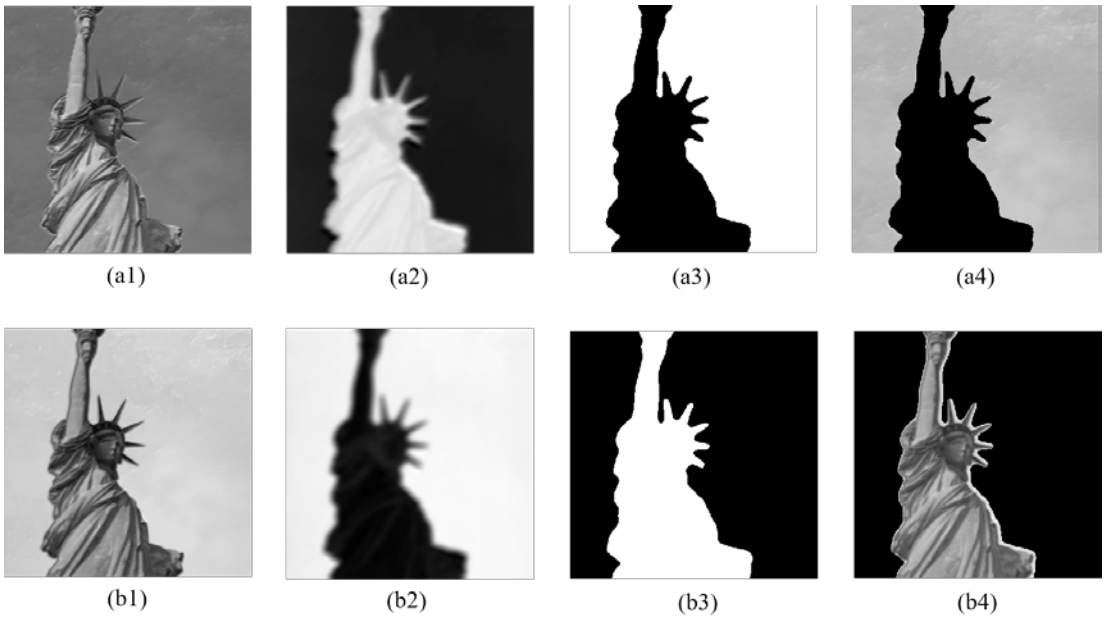


Fig. 7. Crucial features generated by ACoL-JND network.

In summary, the proposed ACoL-JND model utilizes adversarial learning to enable the network to learn complementary information between the handcrafted priors-related regions and non-related regions. The experiment shows that this design can effectively reduce the network's dependence on the accuracy of prior knowledge.

5. Conclusion

In this paper, a novel ACoL-JND model has been proposed for JND estimation. Adversarial complementary learning has been applied to JND modeling, which uses two parallel convolutional networks and a dynamic erase design to extract information from the image. This approach forces two convolutional networks to extract complementary information to solve the problem of information from the non-related handcrafted priors regions being disregarded. Thus, it has comprehensively considered the JND of handcrafted priors-related regions and non-related regions in the image. Besides, to make the handcrafted priors richer, we also have introduced two additional priors that are highly related to JND modeling, including pattern masking (PM) and contrast masking (CM). Eventually, the experimental results have illustrated that our proposed model has outperformed the existing JND models and achieved state-of-the-art performance.

Acknowledgement

Here, we would like to thank the NSF of Shandong Province under Grant ZR2020MF042 and Grant ZR2022MF346, Science and Technology Plan Project of Tangshan Science and Technology Bureau Tangshan Foundation Innovation Team of Digital Media Security under Grant 21130212D for their support of this work.

References

[1] Charles F Hall and Ernest L Hall, "A nonlinear model for the spatial characteristics of the human visual system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 3, pp. 161–170, Mar. 1977. Article (CrossRef Link)

[2] Chun-Man Mak and King Ngi Ngan, "Enhancing compression rate by just-noticeable distortion model for H. 264/AVC," in *Proc. of IEEE International Symposium on Circuits and Systems*, pp. 609–612, Jun. 2009. Article (CrossRef Link)

[3] XK Yang, WS Ling, ZK Lu, Ee Ping Ong, and SS Yao, "Just noticeable distortion model and its applications in video coding," *Signal Processing: Image Communication*, vol. 20, no. 7, pp. 662–680, Aug. 2005. Article (CrossRef Link)

[4] Xinfeng Zhang, Shiqi Wang, Ke Gu, Weisi Lin, Siwei Ma, and Wen Gao, "Just-noticeable difference-based perceptual optimization for JPEG compression," *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 96–100, Jan. 2017. Article (CrossRef Link)

[5] Fanxin Xia, Jian Jin, Lili Meng, Feng Ding, and Huaxiang Zhang, "GAN-Based Image Compression with Improved RDO Process," in *Proc. of International Conference on Image and Graphics*, pp. 361-372, 2023. Article (CrossRef Link)

[6] Feng Ding, Jian Jin, Lili Meng, and Weisi Lin, "JND-Based Perceptual Optimization For Learned Image Compression," *Eprint Arxiv:2302.13092*, Mar. 2023. Article (CrossRef Link)

[7] Weisi 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, May. 2011. Article (CrossRef Link)

[8] Feng Qi, Debin Zhao, Xiaopeng Fan, and Tingting Jiang, "Stereoscopic video quality assessment based on visual attention and just-noticeable difference models," *Signal, Image and Video Processing*, vol. 10, no. 4, pp. 737–744, Apr. 2016. Article (CrossRef Link)

[9] Haiqiang Wang, Ioannis Katsavounidis, Xinfeng Zhang, Chao Yang, and C-C Jay Kuo, "A user model for JND-based video quality assessment: theory and applications," in *Proc. of the SPIE*, 2018. Article (CrossRef Link)

[10] Wen Sun, Jian Jin, and Weisi Lin, "Minimum Noticeable Difference based Adversarial Privacy Preserving Image Generation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 3, pp. 1069–1081, Mar. 2023. Article (CrossRef Link)

[11] Xingxing Zhang, Shupeng Gui, Jian Jin, Zhenfeng Zhu, and Yao Zhao, "ATZSL: Defensive Zero-Shot Recognition in the Presence of Adversaries," *IEEE Transactions on Multimedia*, vol. 26, pp. 15-27, Mar. 2023. Article (CrossRef Link)

[12] Yuan Xue, Jian Jin, Wen Sun, and Weisi Lin, "HVS-inspired adversarial image generation with high perceptual quality," *Journal of Cloud Computing*, vol. 12, no. 1, pp. 1–9, Jun. 2023. Article (CrossRef Link)

[13] Yaqing Niu, Matthew Kyan, Lin Ma, Azeddine Beghdadi, and Sridhar 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, Sep. 2013. Article (CrossRef Link)

[14] Chunxing Wang, Teng Zhang, Wenbo Wan, Xiaoyue Han, and Meiling Xu, "A novel STDM watermarking using visual saliency-based JND model," *Information*, vol. 8, no. 3, pp. 103, Aug. 2017. Article (CrossRef Link)

[15] Sung-Ho Bae and Munchurl Kim, "A novel DCT-based JND model for luminance adaptation effect in DCT frequency," *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 893–896, Sep. 2013. Article (CrossRef Link)

[16] Sung-Ho Bae and Munchurl 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, Aug. 2014. Article (CrossRef Link)

[17] Zhenzhong Chen and Christine Guillemot, "Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 806–819, Jun. 2010. Article (CrossRef Link)
[18] Hadi Hadizadeh, Atiyeh Rajati, and Ivan V Bajić, "Saliency-guided just noticeable distortion estimation using the normalized laplacian pyramid," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1218–1222, Aug. 2017. Article (CrossRef Link)
[19] Yuting Jia, Weisi Lin, and Ashraf A Kassim, "Estimating just-noticeable distortion for video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 820–829, Jul. 2006. Article (CrossRef Link)
[20] Jinjian Wu, Leida Li, Weisheng Dong, Guangming Shi, Weisi Lin, and C-C Jay Kuo, "Enhanced just noticeable difference model for images with pattern complexity," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2682–2693, Jun. 2017. Article (CrossRef Link)
[21] Jinjian Wu, Guangming Shi, Weisi Lin, Anmin Liu, and Fei Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1705–1710, Nov. 2013. Article (CrossRef Link)
[22] Zhipeng Zeng, Huanqiang Zeng, Jing Chen, Jianqing Zhu, Yun Zhang, and Kai-Kuang Ma, "Visual attention guided pixel-wise just noticeable difference model," *IEEE Access*, vol. 7, pp. 132111–132119, Sep. 2019. Article (CrossRef Link)
[23] Jian Jin, Dong Yu, Weisi Lin, Lili Meng, Hao Wang, and Huaxiang Zhang, "Full RGB Just Noticeable Difference (JND) Modelling," *Eprint Arxiv:2203.00629*, Mar. 2022. Article (CrossRef Link)
[24] Lina Jin, Joe Yuchieh Lin, Sudeng Hu, Haiqiang Wang, Ping Wang, Ioannis Katsavounidis, Anne Aaron, and C-C Jay Kuo, "Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis," *Electronic Imaging 2016*, vol. 13, pp. 1–9, 2016. Article (CrossRef Link)
[25] Huanhua Liu, Yun Zhang, Huan Zhang, Chunling Fan, Sam Kwong, C-C Jay Kuo, and Xiaoping Fan, "Deep learning-based picture-wise just noticeable distortion prediction model for image compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 641–656, Aug. 2019. Article (CrossRef Link)
[26] Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeonghoon Park, Shawmin Lei, Xin Zhou, Man-On Pun, Xin Jin, Ronggang Wang, Xu Wang, et al, "VideoSet: A large-scale compressed video quality dataset based on JND measurement," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 292–302, Jul. 2017. Article (CrossRef Link)
[27] Yuhao Wu, Weiping Ji, and Jinjian Wu, "Unsupervised Deep Learning for Just Noticeable Difference Estimation," in *Proc. of IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1-6, 2020. Article (CrossRef Link)
[28] Yun Zhang, Huanhua Liu, You Yang, Xiaoping Fan, Sam Kwong, and CC Jay Kuo, "Deep Learning Based Just Noticeable Difference and Perceptual Quality Prediction Models for Compressed Video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1197–1212, Mar. 2022. Article (CrossRef Link)
[29] Jian Jin, Xingxing Zhang, Xin Fu, Huan Zhang, Weisi Lin, Jian Lou, and Yao Zhao, "Just Noticeable Difference for Deep Machine Vision," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 3452–3461, Jun. 2022. Article (CrossRef Link)
[30] Jian Jin, Yuan Xue, Xingxing Zhang, Lili Meng, Yao Zhao, and Weisi Lin, "HVS-Inspired Signal Degradation Network for Just Noticeable Difference Estimation," *Eprint Arxiv:2208.07583*, Aug. 2022. Article (CrossRef Link)
[31] Xiaolin Zhang, Yunchao Wei, Jiashi Feng, Yi Yang, and Thomas S Huang, "Adversarial complementary learning for weakly supervised object localization," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. Article (CrossRef Link)
[32] Yaxuan Liu, Jian Jin, Yuan Xue, and Weisi Lin, "The First Comprehensive Dataset with Multiple Distortion Types for Visual Just-Noticeable Differences," *Eprint Arxiv:2303.02562*, Mar. 2023. Article (CrossRef Link)

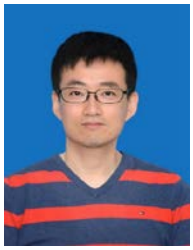
[33] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang, "A simple pooling-based design for real-time salient object detection," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3912-3921, 2019. Article (CrossRef Link)
[34] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. Article (CrossRef Link)
[35] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik, "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 23, no. 2, pp. 684–695, Feb. 2014. Article (CrossRef Link)
[36] Valero Laparra, Johannes Ballé, Alexander Berardino, and Eero P Simoncelli, "Perceptual image quality assessment using a normalized Laplacian pyramid," *Human Vision and Electronic Imaging*, 2016. Article (CrossRef Link)
[37] Zhou Wang, Eero P Simoncelli, and Alan C Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. of The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1398-1402, 2003. Article (CrossRef Link)
[38] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011. Article (CrossRef Link)
[39] Eric Cooper Larson and Damon Michael Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006, Jan. 2010. Article (CrossRef Link)
[40] Hamid R Sheikh and Alan C Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006. Article (CrossRef Link)
[41] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli, "Image quality assessment: Unifying structure and texture similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 5, pp. 2567–2581, May. 2022. Article (CrossRef Link)
[42] Lin Zhang, Ying Shen, and Hongyu Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014. Article (CrossRef Link)
[43] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-595, 2018. Article (CrossRef Link)
[44] Zhou Wang and Eero P Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005. Article (CrossRef Link)
[45] Anmin Liu, Weisi Lin, and Manish Narwaria, "Image quality assessment based on gradient similarity," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012. Article (CrossRef Link)
[46] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009. Article (CrossRef Link)
[47] Tsung-Jung Liu, Weisi Lin, and C-C Jay Kuo, "Image quality assessment using multi-method fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1793–1807, May. 2013. Article (CrossRef Link)
[48] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015. Article (CrossRef Link)
[49] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al., "Rectifier nonlinearities improve neural network acoustic models," in *Proc. of the International Conference on Machine Learning*, 2013. Article (CrossRef Link)

[50] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *Proc. of European Conference on Computer Vision*, pp. 740-755, 2014. Article (CrossRef Link)
[51] Eric C Larson and DM Chandler, "Categorical image quality (CSIQ) database,". [Online]. Available: http://vision.okstate.edu/csiq
[52] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *Eprint Arxiv:1412.6980*, Dec. 2014. Article (CrossRef Link)
[53] Qiuping Jiang, Zhentao Liu, Shiqi Wang, Feng Shao, and Weisi Lin, "Towards Top-Down Just Noticeable Difference Estimation of Natural Images," *IEEE Transactions on Image Processing*, vol. 31, pp. 3697–3712, May. 2022. Article (CrossRef Link)
[54] RECOMMENDATION ITU-R BT, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union*, 2002.

Dong Yu received the M.S. degree in Computer Science and Technology from Shandong Normal University. His research interests include deep learning and image processing.

Jian Jin Member, IEEE. He received the Ph.D. degree from Beijing Jiaotong University. His research interests include image/video/feature compression, visual perceptual modeling, visual quality assessment, and AI security. He was honored with the Excellent Doctoral Dissertation Award of CIE in 2019.

Lili Meng received the Ph.D. degree from Beijing Jiaotong University. She is currently an associate professor at the School of Information Science and Engineering of Shandong Normal University. Her research interests include machine learning and image/video compression.



Zhipeng Chen received the Ph.D. degree in signal and information processing from Beijing Jiaotong University. Now he is an Associate Professor in Tangshan Normal University, China. His research interests include multimedia signal processing, digital forensics, and data hiding.

Huaxiang Zhang received the Ph.D. degree from Shanghai Jiao Tong University. He is currently a professor at the School of Information Science and Engineering of Shandong Normal University. Her research interests include pattern recognition, machine learning, multimodal data retrieval, and person re-identification.