基于熵掩蔽的 DCT 域恰可察觉失真模型

骆琼华，王鸿奎，殷海兵，邢亚芬
（杭州电子科技大学通信工程学院，浙江 杭州 310018）

摘　要：为提高离散余弦变换（discrete cosine transform，DCT）域恰可察觉失真（just noticeable distortion，JND）模型阈值精度并避免跨域操作，将熵掩蔽效应引入 DCT 域 JND 模型。首先，从自由能理论和贝叶斯推理出发，设计基于 DCT 域纹理能量相似性的自回归模型模拟视觉感知过程中的自发预测行为；其次，探索视觉感知与预测残差的映射关系得到块级无序度，并将熵掩蔽效应建模为关于无序度的 JND 阈值调节因子；最后，结合空间对比敏感度函数、亮度自适应掩蔽以及对比度掩蔽，提出基于熵掩蔽的 DCT 域 JND 模型。与现有 DCT 域 JND 模型相比，所提模型所有运算均在 DCT 域执行，更高效简洁。主观、客观实验结果表明，所提模型在感知质量相同或更好的情况下，噪声污染图的平均峰值信噪比（peak signal-to-noise ratio，PSNR）值比其他 4 个 JND 对比模型低 2.04 dB，更符合人眼视觉系统的感知特性。

关键词：恰可察觉失真；人眼视觉系统；熵掩蔽效应；自由能理论；贝叶斯推理
中图分类号：TN91
文献标志码：A
doi: 10.11959/j.issn.1000–0801.2023014

Just noticeable distortion model based on entropy masking in DCT domain

LUO Qionghua, WANG Hongkui, YIN Haibing, XING Yafen
College of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: In order to improve the threshold accuracy of JND (just noticeable distortion) model in DCT (discrete cosine transform) domain and avoid cross-domain operation, entropy masking effect was introduced into DCT-based JND model. Firstly, starting from the free-energy theory and the Bayesian inference, an autoregressive model based on texture-energy similarity in DCT domain was designed to simulate the spontaneous prediction behavior of visual perception. Secondly, the mapping relationship between visual perception and prediction residuals were explored to obtain the disorder intensity in block level. Thirdly, the entropy masking effect was modeled as a JND threshold modulation factor of disorder intensity. Finally, the JND model in DCT domain for the entropy masking was proposed by fusing the contrast sensitivity function, the luminance adaptive masking, the contrast masking. Compared with the

收稿日期：2022–08–03；**修回日期：**2023–01–09
基金项目：国家自然科学基金资助项目（No.62202134，No.61972123，No.61931008，No.62031009）；浙江省“尖兵”“领雁”研发攻关计划项目（No.2023C01149，No.2022C01068）
Foundation Items: The National Natural Science Foundation of China (No.62202134, No.61972123, No.61931008, No.62031009), Zhejiang Provincial “Pioneer” and “Leading Goose” Research and Development Project (No.2023C01149, No.2022C01068)

existing JND model in DCT domain, the proposed model performed all operations in DCT domain, which was more efficient and concise. The subjective and objective experimental results indicate that the proposed JND model shows greater tolerance to distortion with better perceptual quality.

Key words: JND, human visual system, entropy masking effect, free-energy theory, Bayesian inference

0 引言

人类视觉系统（human visual system，HVS）由于其潜在的生理和心理机制无法感知一定阈值以下的图像失真，这一阈值被称为恰可察觉失真（just noticeable distortion，JND）阈值[1]。JND 模型揭示了人类视觉系统对图像失真的感知特性，被广泛应用于图像和视频处理中，如图像和视频压缩/编码[2]、质量评价[3]、数字水印[4]和图像增强[5]等。

过去二十年里，JND 阈值估计取得了巨大进展。现有 JND 估计模型根据计算域可分为两类：像素域 JND 模型和子带域 JND 模型。其中，像素域 JND 模型直接计算图像和视频中每个像素的 JND 阈值[6]；子带域 JND 模型主要在压缩域中估计 JND 阈值，如小波域和 DCT 域[7]。本文重点关注基于 DCT 域的 JND 阈值估计，因为 DCT 域 JND 模型可直接应用于基于 DCT 变换的图像/视频压缩领域[2]。

DCT 域 JND 模型主要由对比敏感度函数（contrast sensitivity function，CSF）、亮度自适应（luminance adaptive，LA）掩蔽效应和对比度掩蔽（contrast masking，CM）效应构成[7]。纵观现有 JND 模型，其核心问题是如何准确高效地估计 CM 模型。DCT 域 CM 模型通常将图像块分为 3 种类型（即平面、边缘和纹理），并为这 3 种类型设置 3 个不同的权重以突出纹理区域[7]。然而内部生成机制（internal generation mechanism，IGM）对有序纹理内容是自适应的，因此这些有序纹理区域会被高估。

实际上，HVS 并非逐字翻译输入场景，而是通过 IGM 主动推断输入场景[8]。自由能理论为人类的行为、感知和学习提供了统一的解释，其基本思想是所有的适应性生物制剂都能抵抗紊乱的自然趋势[9]。HVS 在自由能理论的指导下试图处理尽可能多的结构信息而避免不确定信息。这一特性揭示了 HVS 的感知局限性，包含大量不确定信息的无序纹理区域的 JND 阈值往往相对较高。近年来，自由能理论已被引入像素域 JND 模型中进行有意义的探索[10]。Wu 等[10]通过贝叶斯预测模拟视觉系统自发的预测行为，将图像分为有序成分和无序成分进行独立处理，以熵掩蔽（entropy masking，EM）的形式引入像素域 JND 阈值估计。在此基础上，Wu 等[11]引入模式掩蔽取代熵掩蔽和对比度掩蔽，并提出一个增强型的像素域 JND 模型。Zeng 等[12]利用相对全变分模型和方向复杂度，将图像分成结构图像、有序纹理图像和无序纹理图像，并结合视觉显著性构建像素域 JND 模型。Liu 等[13]将屏幕内容图像分解为屏幕内容集和非屏幕内容集进行处理，并研究了边缘掩蔽和模式掩蔽的组合掩蔽效应。Wang 等[14]根据分层预测编码理论将视觉感知分为 3 个阶段，并结合正负感知效应提出一个新的像素域 JND 模型。

基于上述自由能理论在像素域 JND 阈值估计中的探索，本文认为在 DCT 域 JND 阈值估计中合理引入自由能理论与熵掩蔽效应将有助于提高 DCT 域 JND 模型的准确性。由于大多数图像/视频资源以压缩形式进行有效存储和传输，直接对压缩数据而非其解压缩版本进行操作变得非常重要。本文拟在避免跨域运算的前提下，引入熵掩蔽效应并提出 DCT 域 JND 建模方法。

1 DCT 域 JND 模型概述

1.1 DCT 域 JND 模型框架

本文主要研究基于 DCT 域的 JND 阈值估计。对于 $N\times N$ 大小的 DCT 块，第(i,j)个 DCT 系数计算式如下。

$$X_{i,j}=\lambda(i)\lambda(i)\times\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}I(k,l)\cos\left[\frac{(2k+1)i\pi}{2N}\right]\cos\left[\frac{(2l+1)j\pi}{2N}\right] \tag{1}$$

其中，$I(k,l)$ 为第 (k,l) 个图像像素，$k,l=0,1,\cdots,N-1$，$N=8$。$\lambda(i)$ 和 $\lambda(j)$ 为补偿系数，计算如下。

$$\lambda(i)=\lambda(j)=\begin{cases}\sqrt{1/N}, & i,j=0\\ \sqrt{2/N}, & i,j=1,2,\cdots,N-1\end{cases} \tag{2}$$

通常，DCT 域 JND 模型被表示为多个调节因子的乘积，这源于 Watson[15]的 DCTune 模型。本文提出的 JND 模型主要考虑 4 种掩蔽效应，计算如下。

$$J=J_{\text{base}}\times J_{\text{LA}}\times J_{\text{CM}}\times J_{\text{EM}} \tag{3}$$

其中，J_{base} 即 CSF-JND，主要基于空间对比敏感度函数进行估计，不包括其他掩蔽效应，这里采用 Wei[16]的 J_{base} 模型；J_{LA} 对应亮度自适应掩蔽效应的 JND 阈值调节因子，主要描述平均背景亮度与视觉阈值之间的关系，这里采用 Wei[16]的分段线性函数；J_{CM} 代表对比度掩蔽效应的 JND 阈值调节因子，主要反映当前视觉信号在其他视觉信号存在时的可见度衰减效应，本文采用 Bae[7]的 CM 模型，并在第 1.2 节中对其改进；J_{EM} 为本文在 DCT 域提出的熵掩蔽效应的 JND 阈值调节因子，在第 2 节中进行详细描述。

1.2　CM 模型改进

CM 模型主要受对比度强度影响。Bae[7]将对比度强度 $\upsilon(n)$ 定义为 DCT 块的标准差与归一化功率谱密度峰度之比，其第 n 个 DCT 块的对比度强度计算如下。

$$\tau(n)=\frac{2^{1.4}\left(\sum_{\omega\in\gamma,\omega\neq0}\omega^2X^2(\omega)\right)^2}{K^{1.4}N^{2.1}\sum_{\omega\in\gamma,\omega\neq0}\omega^4X^2(\omega)} \tag{4}$$

其中，K 是像素强度的动态范围（8 位图像格式为 255），γ 表示 $N\times N$ 的 DCT 块区域，$X(\omega)$ 是频率 ω 下的 DCT 系数值，计算同 Bae[7]的模型。

然而边缘区域的标准差普遍大于纹理区域的标准差，这并不符合 HVS 对边缘区域敏感的特性。有实验表明[17]，即使两个图像在局部区域中具有相似的功率谱、背景亮度水平和能量含量，纹理图像在不损害感知质量的情况下隐藏的噪声是边缘图像的 3 倍。因此，本文利用 DCT 域的纹理能量（texture energy，TE）块分类法[18]将图像分为边缘块、纹理块和平坦块，TE 块分类如图 1 所示。分类结果如图 1（b）所示，白色、灰色和黑色分别代表边缘块、纹理块和平坦块。根据块分类结果，本文引入抑制因子 $\upsilon(n)$ 抑制边缘块和平坦块的对比度强度，于是对比度强度被表示为 $\upsilon(n)\cdot\tau(n)$。对于边缘块、纹理块和平坦块，$\upsilon(n)$ 分别设置为 1/3、1 和 1/3[17]。

根据 Bae[7]的实验，CM-JND 阈值随着对比度强度线性增加。因此，J_{CM} 被建模为对比度强度的线性函数，具体如下。

$$J_{\text{CM}}(n,\omega_{i,j})=\upsilon(n)\cdot\tau(n)\cdot f(\omega_{i,j})+1 \tag{5}$$

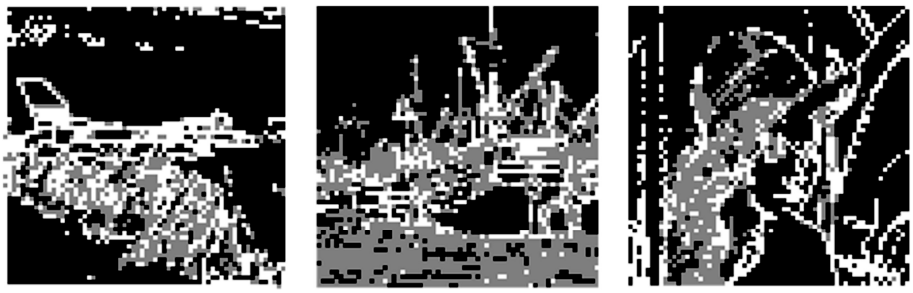
其中，$f(\omega_{i,j})$ 是频率线性函数梯度，计算同 Bae[7]的模型。引入抑制因子 $\upsilon(n)$ 前后对比如图 2 所示，截取边缘区域（即黑色方框区域）进行更清晰的对比，通过引入抑制因子 $\upsilon(n)$，边缘块的对比度强度被有效抑制，边缘区域的图像质量也得到明显改善。

2　基于熵掩蔽的 JND 阈值调节因子

2.1　贝叶斯预测模型

HVS 在自由能理论的指导下可以准确预测并充分理解有序内容，同时粗略感知无序内容[9]。显然，与无序区域相比，HVS 对有序区域更敏感，其 JND 阈值更小。由于贝叶斯推理是信息预测的有力工具，本文采用贝叶斯大脑理论[19]模拟人脑中的 IGM 进行有序信息预测，并建立 DCT 域自回归预测模型。贝叶斯大脑理论的基本原理是，大脑有一个试图以最小误差概率表示传感器信息的模型[19]。对于图像而言，贝叶斯大脑系统以最大化条件概率表示像素值。

(a) 原图



(b) 块分类结果图
(白色、黑色和灰色分别代表边缘块、纹理块和平坦块)

图 1　TE 块分类

(a) 抑制前的对比度强度图　(b) 抑制前的注噪图　(c) 抑制前的注噪边缘图

(d) 抑制后的对比度强度图　(e) 抑制后的注噪图　(f) 抑制后的注噪边缘图

图 2　引入抑制因子前后对比

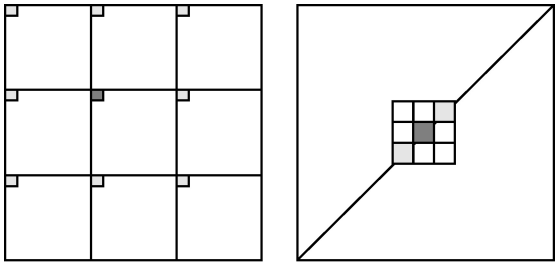
然而，$N\times N$ 的 DCT 变换割裂了图像块之间的相关性，因此，本文预测模型以当前块为中心，选取其邻近一圈的 8 个块（$\chi=\{B_1,\cdots,B_8\}$）的同位 DCT 系数对中心块的 DCT 系数进行预测，DCT 域贝叶斯预测示例如图 3 所示。例如，中心块(0, 0)位置的 DCT 系数由周围 8 个块(0, 0)位置的 DCT 系数共同预测得到，如图 3（a）所示。对于 $N=8$ 的 DCT 块，Zhang[18]令 L、M 和 H 分别表示低频

（low frequency，LF）、中频（medium frequency，MF）和高频（high frequency，HF）组中的绝对 DCT 系数值之和，8×8 DCT 块能量分布如图 4 所示。DCT 块的纹理能量 E_{tex} 可用 M 和 H 计算，因而也可得到中心块与其周围块间的纹理能量差 ΔE_{tex}，计算如下。

$$E_{\text{tex}} = M + H \tag{6}$$

$$\Delta E_{\text{tex}} = | E_{\text{ctex}} - E_{\text{stex}} | \tag{7}$$

其中，E_{ctex} 和 E_{stex} 分别为中心块和周围块的纹理能量。



(a) 预测示例　　(b) 边缘区域预测示例

图 3　DCT 域贝叶斯预测示例

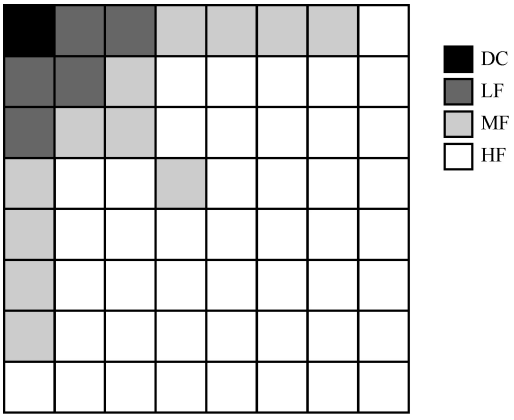


图 4　8×8 DCT 块能量分布

若中心块与周围块非常相似，几乎没有不确定性，则可用周围块准确推断出当前块，即 ΔE_{tex} 接近或等于 0；反之，中心块与周围块的相似度越低，则 ΔE_{tex} 越大。因此，ΔE_{tex} 一定程度上可以反映图像内容的有序程度，且二者呈负相关。本文利用 ΔE_{tex} 构建反比例函数度量 DCT 块间的相似度。由于存在 $\Delta E_{\text{tex}} = 0$ 的情况，且指数函数能更有效地根据 ΔE_{tex} 的大小分配相似度，最终相似度计算如下。

$$S = 1 / \exp(\Delta E_{\text{tex}}) \tag{8}$$

通过模仿 IGM 的推理机制，本文试图在 DCT 域创建一个预测模型，该模型根据周围 DCT 块的 DCT 系数及其相似度预测当前块的 DCT 系数。中心块 B_{c} 与其邻域 χ 的相似度 $S(B_{\text{c}};\chi)$ 越高，则图像内容的无序程度越低；并且周围块 B_z 的相似度 $S(B_{\text{c}};B_z)$ 越大，在 IGM 预测过程中的作用越大。因此本文将中心块与其周围块的相似度 $S(B_{\text{c}};B_z)$ 作为自回归系数，建立 DCT 域贝叶斯预测模型，如式（9）所示。

$$P'(X) = \sum_{B_z \in \chi} \frac{S(B_{\text{c}};B_z)}{\sum_{\text{u}} S(B_{\text{c}};B_{\text{u}})} \cdot P(X) + \varepsilon \tag{9}$$

其中，P 为 DCT 输入图像，P' 为 DCT 预测图像。$\frac{S(B_{\text{c}};B_z)}{\sum_{\text{u}} S(B_{\text{c}};B_{\text{u}})}$ 为归一化的相似度，ε 为 ±1 的双极性随机噪声。

对于边缘区域，HVS 应能自适应地进行预测。如图 3（b）所示，中心块位于平坦区域的一条直线（即边缘区域）上时，应当只由同处边缘区域的周围块进行预测。因此，本文采用 DCT 域的 TE 块分类法[18]将图像分为边缘块和非边缘块进行预测。当中心块为边缘块时，仅由同为边缘块的周围块进行相似度加权预测；当中心块为非边缘块时，仅由非边缘块进行相似度加权预测；而当中心块为边缘块（非边缘块）且其周围块均为非边缘块（边缘块）时，中心块由周围 8 个块共同预测得到。

DCT 域贝叶斯预测结果如图 5 所示，本文获得的预测图像和残差图像存在明显的块效应，这是由 8×8 的 DCT 所引起的。为证明本文预测模型的准确性，采用多尺度结构相似性（multi-scale structural similarity，MS-SSIM）指数度量预测图像与原始图像之间的相关性，高于 0.5 的 MS-SSIM 指数值表明图像间具有很强的相似性[20]。从图 5（b）及其对应的 MS-SSIM 指数值可以看到，本文提出的 DCT 域贝叶斯预测模型能较为准确地预测图像内容。

图 5　DCT 域贝叶斯预测结果

2.2　熵掩蔽模型

根据IGM理论，大脑作为一个主动推理系统，能准确地预测有序内容，避免无序信息[9]。因此本文将原始 DCT 图像 P 与预测 DCT 图像 P' 之间的差视为残差图像 R，如图 5（c）所示。

$$R = |P - P'| \tag{10}$$

由于DCT后绝大部分能量都集中在左上角的低频系数中，DCT 系数之间幅度相差很大。而DCT系数幅度越大，其预测残差往往越大。因此这里将 DCT 系数残差与原始系数值相除，并对其相除结果进行归一化处理得到无序度块。第 n 个 DCT 块的无序度由其无序度块求平均得到，计算如下。

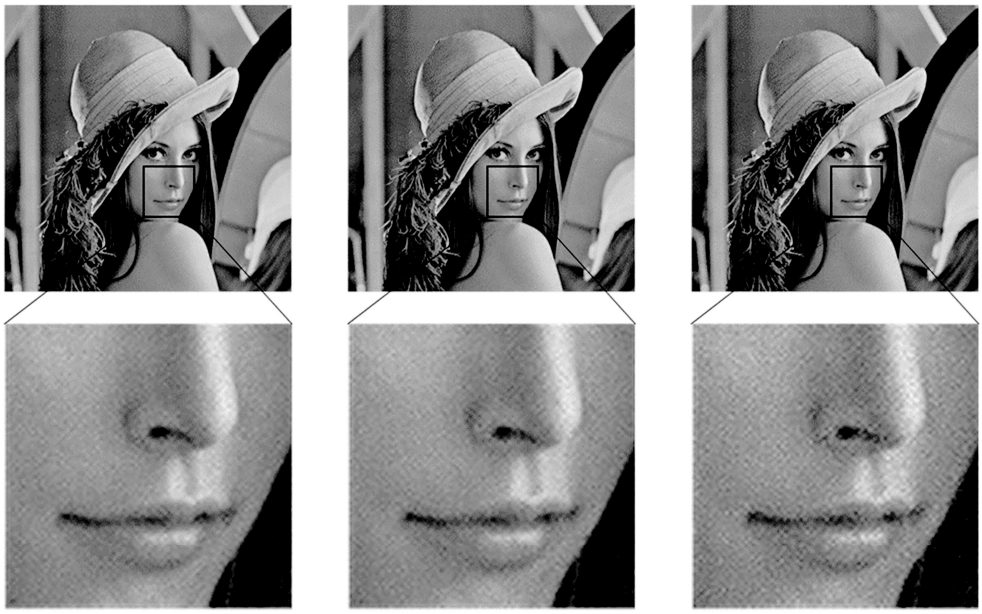
$$\zeta(n) = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\mathbb{N}\left(\left|\frac{R(X_{i,j})}{P(X_{i,j})}\right|\right)}{N \times N} \tag{11}$$

其中，$\mathbb{N}$ 是归一化运算。$X_{i,j}$ 为 DCT 块中第 (i,j) 个 DCT 系数。

基于 DCT 域的 JND 模型一般表示为多个调节因子的乘积形式，因此不存在熵掩蔽的有序纹理区域的 J_{EM} 可表示为 1，而存在熵掩蔽的无序纹理区域的 J_{EM} 应大于 1。无序度 $\zeta(n)$ 代表输入图像内容的不确定性，$\zeta(n)$ 越大，则输入图像包含的不确定信息越多，可隐藏的噪声越多。由于指数函数能将更多的噪声分配给 $\zeta(n)$ 较大的图像块，因此本文将 J_{EM} 构建为：

$$J_{\mathrm{EM}}\left(\zeta(n)\right) = 1 + \alpha \cdot \exp\left(\zeta(n)\right) \tag{12}$$

其中，α 是比例因子，根据主观观看结果，按照经验将其设为 0.15。这里选取“IVC”数据集[22]中的一幅图像，给出主观质量对比的例子，不同 α 值得到的主观质量对比如图 6 所示。从主观质量的角度可以看到，当 α 在取 0.1 和 0.15 时图像主观质量相近，而 $\alpha = 0.2$ 时主观质量明显受损；从噪声隐藏能力的角度，可以看到 α 依次取 0.15 和 0.2 时，PSNR 值分别降低 0.71 dB 和 0.27 dB。综



(a) α=0.1, PSNR=27.43 dB　(b) α=0.15, PSNR=26.72 dB　(c) α=0.2, PSNR=26.45 dB

图 6　不同 α 值得到的主观质量对比

上所述，α 取 0.15 时可以在不影响图像质量的情况下降低较多的 PSNR 值。

3　实验结果与分析

3.1　对比实验

为评估所提出 JND 模型的性能，本文在 3 个数据集（IVC[21]、TID2013[22]和 LabelMe[23]）中选择具有不同分辨率的 30 幅图像作为实验样本[24]。其中，IVC 和 TID2013 是用于评估 JND 模型性能的经典图像数据集，分辨率分别为 512 dpi×512 dpi 和 512 dpi×384 dpi。此外，在 LabelMe 数据集中，分辨率为 1 024 dpi×768 dpi 的 10 幅图像被用来评估 JND 模型在高分辨率图像上的性能。本文将这些图像分别命名为“I_{01}”～“I_{10}”、“T_{01}”～“T_{10}”和“L_{01}”～“L_{10}”。

本文提出的 JND 估计模型是在 DCT 域中结合 EM 效应而构建的。因此为验证本文 DCT 域 JND 模型的有效性，这里选择研究工作相近的 3 个 JND 模型进行对比实验，分别命名为 Zeng2019[12]、Liu2020[13]以及 Li2022[25]。其中，Zeng2019 和 Liu2020 是考虑自由能理论的两种像素域 JND 模型，而 Li2022 是融合边缘掩蔽与中心凹掩蔽的像素域 JND 模型。

通常在更精确 JND 模型的指导下，图像能保持感知质量不变并隐藏更多噪声。在基于 DCT 域的 JND 估计中，噪声被添加到图像中的每个 DCT 系数[24]，如式（13）所示。

$$X'(n,i,j)=X(n,i,j)+\mathrm{rand}(n,i,j)\cdot J(n,i,j) \tag{13}$$

其中，$X(n,i,j)$ 是第 n 个 DCT 块中第 (i,j) 个 DCT 系数，而 $X'(n,i,j)$ 是注入 JND 噪声的 DCT 系数。$\mathrm{rand}(n,i,j)$ 是大小为 ±1 的双极性随机噪声。

JND 模型的噪声容忍能力是用峰值信噪比（peak signal-to-noise ratio，PSNR）衡量的，即相同感知质量下 PSNR 越低，JND 模型估计越准确。8 位灰度图像的 PSNR 由原始图像和噪声注入图像之间的均方误差（mean squared error，MSE）计算得到，如式（14）所示。

$$\mathrm{PSNR}=10\cdot\lg\left(\frac{(2^8-1)^2}{\mathrm{MSE}}\right) \tag{14}$$

由于图像的最终接收者为人眼，仅以 PSNR 评估 JND 模型性能不够准确，因此采用 MS-SSIM 指数和平均意见分数（mean opinion score，MOS）评价图像的主观质量。其中，MS-SSIM 指数通过

图像间结构相似性评估图像质量，是一种更适合HVS 的感知质量评价指标[26]。本文根据 ITU-R BT.500-11 标准展开主观质量评估（subjective quality assessment，SQA）实验[24]，该实验用于测量原始图像和噪声注入图像间的主观质量差异。每次测试中，为了更好地比较主观质量，将原始图像作为标准放在左边，4 个注入噪声的图像（本文模型和用于对比的 3 个模型）以随机顺序在右边并列显示。主观评价标准和分数被分为 4 个等级，以指示噪声注入图像相对于原始图像的失真程度。评价标准具体为：“0”表示右边图像质量没有下降，“–1”表示轻微下降，“–2”表示下降较多，“–3”表示下降严重。25 位视力正常或矫正视力正常的受试者被邀请评估噪声注入图像的感知质量，个人评价分数以 MOS 的形式提供。在 SQA 实验中，显示器的分辨率为 1 920 dpi×1 080 dpi，并将观看距离设置为显示屏高度的 4 倍[24]。

3.2 性能评估

JND 模型的性能是以 PSNR、MS-SSIM 指数和 MOS 衡量的。被测 JND 模型的良好性能意味着由 JND 模型注入噪声的图像与其他模型相比具有更低的 PSNR 值以及更高的 MS-SSIM 指数值和 MOS 值。表 1～表 3 展示了注入相同噪声后本文模型与其他 3 个模型在不同图像数据集上的 PSNR、MS-SSIM 指数和 MOS 值，其中由于 MS-SSIM 指数值相差较小，在这里保留 3 位小数。为了更直观地对比 JND 模型性能，采用柱状图的形式描述不同 JND 模型在 3 个数据集上主客观评价指标的平均值，如图 7 所示。本文以 PSNR 测量 JND 模型隐藏噪声的能力。由表 1～表 3 可知，本文模型在 3 个分辨率不同的数据集上均具有最低的 PSNR 值。具体而言，本文模型在 3 个数据集上获得的平均 PSNR 值为 26.75 dB，分别比其他模型低 2.52 dB、2.12 dB 和 1.47 dB，如图 7（a）所示。此外，本文以 MS-SSIM 指数和 MOS 测量噪声注入图像的感知质量。由表 1～表 3 可知，本文模型在 3 个数据集上均获得最高的 MS-SSIM 指数值和 MOS 值。具体来说，本文在 3 个数据集上获得的平均 MS-SSIM 指数值为 0.963，分别比其他 3 个模型高 0.020、0.018 和 0.023，如图 7（b）所示；而平均 MOS 值为–0.29，比其他模型高出 0.17、0.13 和 0.16，如图 7（c）所示。由此可见，与其他

表 1 不同 JND 模型在“IVC”数据集上的性能对比

IVC	PSNR				MS-SSIM 指数				MOS			
	Zeng2019	Liu2020	Li2022	本文	Zeng2019	Liu2020	Li2022	本文	Zeng2019	Liu2020	Li2022	本文
I_{01}	29.80	30.92	31.44	26.19	0.947	0.962	0.977	0.959	–0.61	–0.38	–0.39	–0.20
I_{02}	29.25	28.39	27.79	25.26	0.964	0.966	0.962	0.968	–0.52	–0.50	–0.38	–0.25
I_{03}	30.15	30.02	30.52	25.98	0.961	0.964	0.973	0.962	–0.56	–0.50	–0.36	–0.20
I_{04}	27.85	27.65	25.93	26.04	0.945	0.944	0.930	0.968	–0.52	–0.47	–0.42	–0.29
I_{05}	28.20	28.56	26.77	26.61	0.956	0.963	0.956	0.969	–0.47	–0.37	–0.40	–0.24
I_{06}	30.01	29.42	30.38	25.77	0.964	0.971	0.981	0.966	–0.46	–0.41	–0.47	–0.30
I_{07}	29.31	30.16	29.49	26.67	0.956	0.967	0.968	0.967	–0.54	–0.46	–0.31	–0.19
I_{08}	29.65	30.48	28.86	26.72	0.957	0.968	0.963	0.980	–0.54	–0.46	–0.36	–0.28
I_{09}	30.14	26.22	28.98	25.24	0.982	0.975	0.984	0.964	–0.29	–0.43	–0.28	–0.22
I_{10}	28.90	28.73	27.99	26.86	0.949	0.959	0.951	0.964	–0.41	–0.45	–0.36	–0.23
平均值	29.33	29.05	28.81	**26.13**	0.958	0.964	0.965	**0.967**	–0.49	–0.44	–0.37	**–0.24**

表 2　不同 JND 模型在“TID2013”数据集上的性能对比

TID2013	PSNR				MS-SSIM 指数				MOS			
	Zeng2019	Liu2020	Li2022	本文	Zeng2019	Liu2020	Li2022	本文	Zeng2019	Liu2020	Li2022	本文
T_{01}	29.50	29.24	27.72	26.83	0.932	0.932	0.910	0.954	–0.50	–0.39	–0.54	–0.25
T_{02}	28.92	29.31	27.40	27.12	0.934	0.941	0.921	0.959	–0.43	–0.28	–0.54	–0.30
T_{03}	29.42	29.27	27.43	26.27	0.972	0.974	0.965	0.978	–0.41	–0.37	–0.46	–0.31
T_{04}	28.37	26.53	26.41	24.56	0.974	0.975	0.968	0.972	–0.40	–0.43	–0.37	–0.27
T_{05}	29.91	29.73	29.20	26.24	0.954	0.963	0.962	0.961	–0.46	–0.38	–0.49	–0.30
T_{06}	30.08	28.73	29.37	26.44	0.961	0.969	0.971	0.963	–0.64	–0.40	–0.51	–0.35
T_{07}	27.12	26.84	25.18	26.65	0.904	0.903	0.874	0.954	–0.49	–0.28	–0.51	–0.28
T_{08}	30.47	30.28	29.19	26.52	0.957	0.962	0.955	0.958	–0.43	–0.36	–0.45	–0.28
T_{09}	30.01	28.11	28.26	25.73	0.972	0.974	0.976	0.967	–0.44	–0.37	–0.43	–0.33
T_{10}	30.15	30.17	29.37	26.91	0.950	0.955	0.950	0.958	–0.44	–0.27	–0.47	–0.27
平均值	29.40	28.82	27.95	**26.33**	0.951	0.955	0.945	**0.962**	–0.46	–0.35	–0.48	**–0.29**

表 3　不同 JND 模型在“LabelMe”数据集上的性能对比

LabelMe	PSNR				MS-SSIM 指数				MOS			
	Zeng2019	Liu2020	Li2022	本文	Zeng2019	Liu2020	Li2022	本文	Zeng2019	Liu2020	Li2022	本文
L_{01}	31.88	31.21	32.78	29.68	0.946	0.943	0.966	0.980	–0.45	–0.35	–0.56	–0.40
L_{02}	27.25	26.95	25.82	27.21	0.883	0.877	0.866	0.946	–0.26	–0.55	–0.52	–0.35
L_{03}	30.57	30.86	29.56	28.50	0.951	0.952	0.957	0.976	–0.43	–0.45	–0.46	–0.42
L_{04}	29.27	28.80	27.96	27.23	0.945	0.943	0.946	0.967	–0.57	–0.45	–0.48	–0.34
L_{05}	29.13	28.55	27.63	27.88	0.935	0.932	0.924	0.964	–0.47	–0.49	–0.47	–0.34
L_{06}	25.72	24.84	23.96	27.50	0.841	0.830	0.810	0.947	–0.4	–0.47	–0.50	–0.42
L_{07}	31.33	31.25	30.41	28.92	0.957	0.955	0.960	0.970	–0.45	–0.40	–0.42	–0.29
L_{08}	30.35	30.12	29.22	27.14	0.959	0.960	0.959	0.960	–0.5	–0.56	–0.63	–0.32
L_{09}	28.68	28.47	26.89	27.09	0.914	0.914	0.891	0.947	–0.42	–0.53	–0.46	–0.33
L_{10}	26.71	26.26	24.91	26.90	0.872	0.865	0.839	0.939	–0.44	–0.57	–0.59	–0.27
平均值	29.09	28.73	27.91	**27.80**	0.920	0.917	0.912	**0.960**	–0.44	–0.48	–0.51	**–0.35**

注：加粗字体为最优结果。

3 个模型相比，本文提出的 DCT 域 JND 模型更精确，能隐藏更多噪声，获得更好的图像感知质量。

为进一步验证本文 JND 模型的性能，这里给出一些主观质量比较的例子。本文从 3 个分辨率不同的图像数据集中分别选取一张图片（“I_{08}”“T_{03}”“L_{07}”）作为示例，并截取局部图像（即黑色方框区域）进行更清晰的对比，如图 8～图 10 所示。图 8 截取部分为边缘平滑且亮度差较大的柱子区域，该区域包含的不确定信息很少，其失真极容易被人眼察觉，因此本文提出的 JND 模型在此区域注入少量噪声，如图 8（d）所示。而 Zeng2019、Liu2020 和 Li2022 的 JND 模型高估了

此区域的掩蔽效应，注入噪声过量，造成明显的图像失真，且 PSNR 值分别比本文模型高 2.93 dB、3.76 dB 和 2.14 dB，如图 8（a）、图 8（b）和图 8（c）所示。图 9 截取部分为包含有序纹理的窗户区域，人眼能主动预测此区域纹理，对此区域的噪声比较敏感，因此本文提出的 JND 模型在此区域注入适量噪声，如图 9（d）所示。Li2022 的 JND 模型虽也获得了较低的 PSNR 值，但其忽略了人眼对有序区域的敏感度，在此区域注入过量噪声，引起图像较大失真，如图 9（c）所示。Zeng2019 和 Liu2020 在此区域注入噪声稍有过量，获得的图像感知质量不佳，且 PSNR 值分别比本文模型高 3.15 dB 和 3.00 dB，如图 9（a）、图 9（b）所示。图 10 截取部分为包含无序纹理的树枝区域，此区域的失真不容易被察觉，因此本文提出的 JND 模型在此区域注入大量噪声，在不影响图像感知质量的前提下获得最低的 PSNR 值，如图 10（d）所示。而其他 3 个 JND 模型在此区域注入的噪声均稍有过量，造成轻微图像失真，且 PSNR 值分别比本文模型高 2.41 dB、2.33 dB 和 1.49 dB，如图 10（a）、图 10（b）和图 10（c）所示。

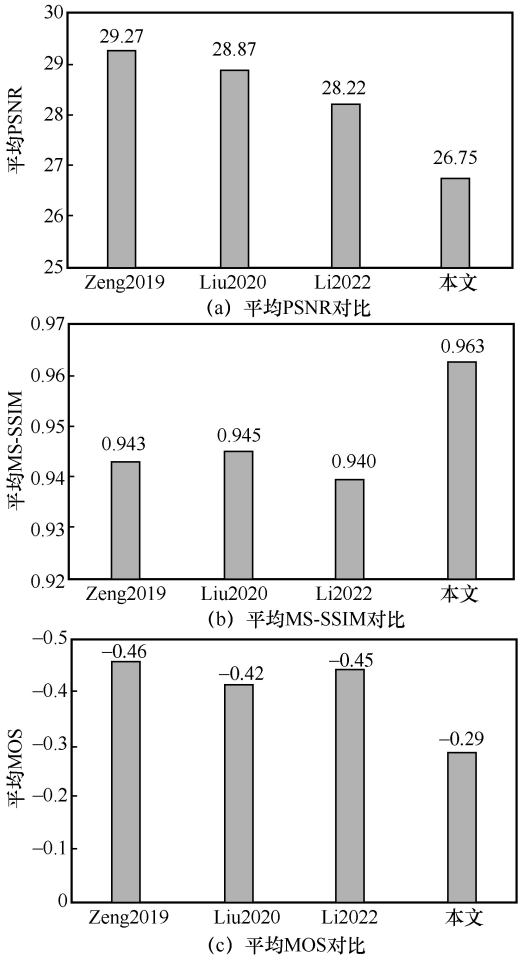
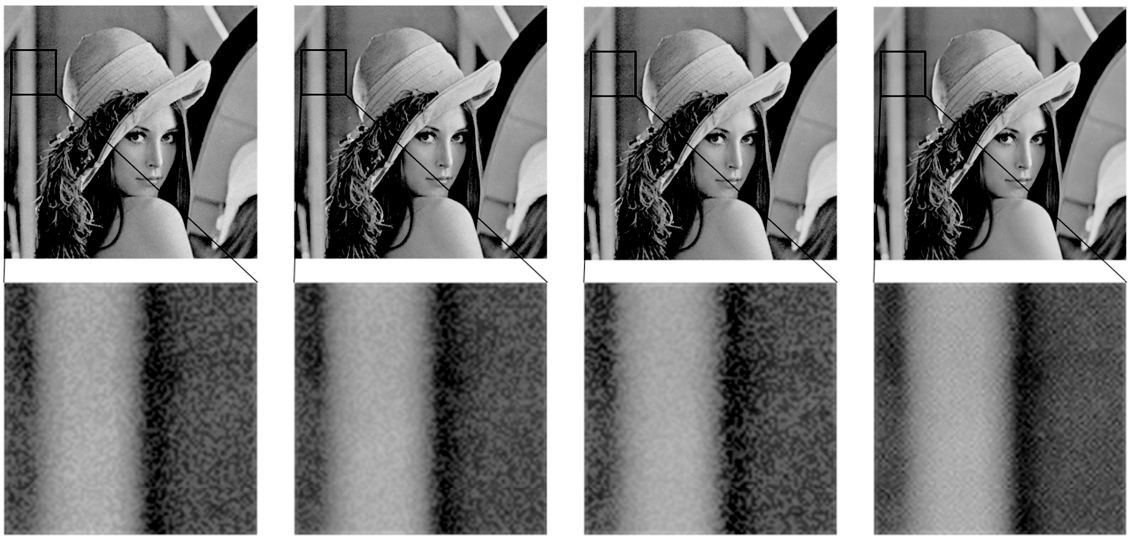
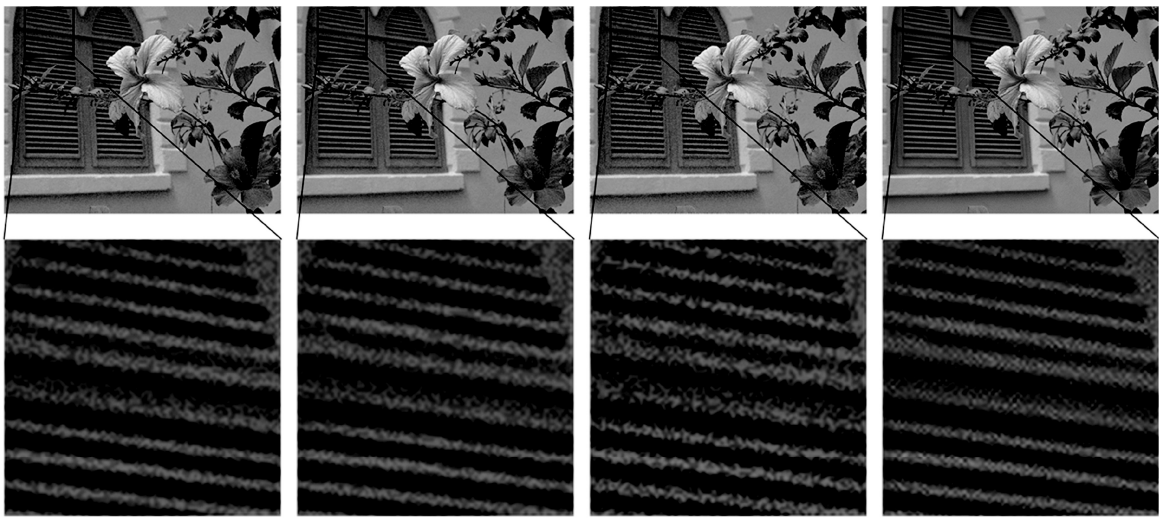


图 7　不同 JND 模型在 3 个数据集上的主客观指标平均值对比



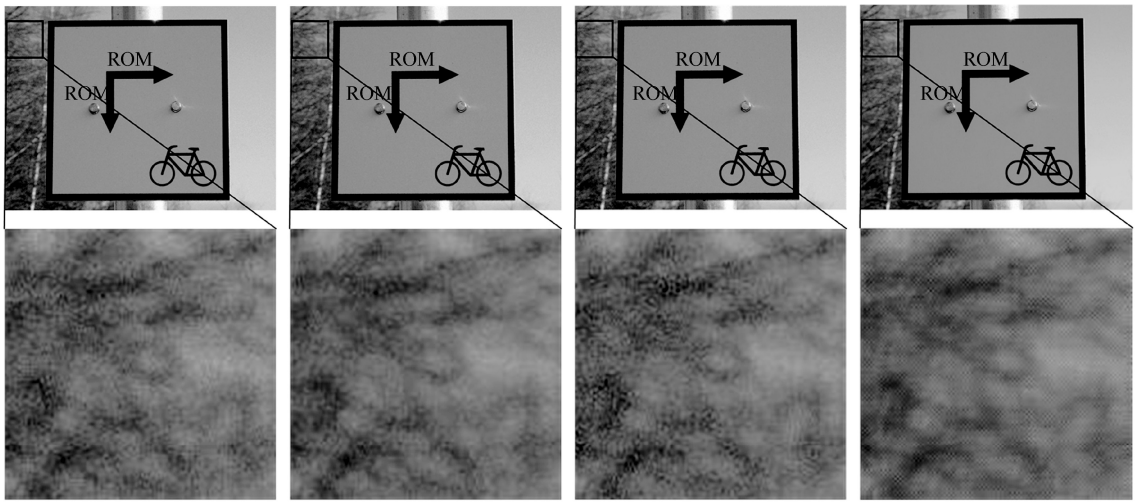
(a) Zeng2019, PSNR=29.65 dB　(b) Liu2020, PSNR=30.48 dB　(c) Li2022, PSNR=28.86 dB　(d) 本文, PSNR=26.72 dB

图 8　不同 JND 模型在“ I_{08} ”图像上的对比



（a）Zeng2019, PSNR=29.42 dB　（b）Liu2020, PSNR=29.27 dB　（c）Li2022, PSNR=27.43 dB　（d）本文, PSNR=26.27 dB

图 9　不同 JND 模型在“T_{03}”图像上的对比



（a）Zeng2019, PSNR=31.33 dB　（b）Liu2020, PSNR=31.25 dB　（c）Li2022, PSNR=30.41 dB　（d）本文, PSNR=28.92 dB

图 10　不同 JND 模型在“L_{07}”图像上的对比

以上主观、客观实验证明，本文提出的 JND 模型能充分考虑人眼视觉特性，在平坦、有序纹理、无序纹理区域分别注入少量、适量、大量噪声，实现噪声的合理分配。综上所述，本文模型能在注入相同噪声的情况下获得最好的感知质量和最低的 PSNR 值，性能优于其他 3 个 JND 模型。

4　结束语

本文通过探索 HVS 主动预测输入信号的特性，提出基于熵掩蔽的 DCT 域 JND 模型。算法充分考虑 DCT 块间的相关性，设计基于纹理能量相似性的 DCT 域贝叶斯预测模型，利用预测残差计算块级无序度描述图像内容的不确定性；最终将熵掩蔽效应构建为关于无序度的调节因子，以指示更多噪声分配到包含大量不确定信息的无序纹理区域。主观、客观实验结果表明，所提模型能在避免跨域操作的同时去除更多感知冗余，性能优于现有 JND 模型。

参考文献：

[1] JAYANT N, JOHNSTON J, SAFRANEK R. Signal compression based on models of human perception[J]. Proceedings of the IEEE, 1993, 81(10): 1385-1422.

[2] KI S, DO J, KIM M. Learning-based JND-directed HDR video

preprocessing for perceptually lossless compression with HEVC[J]. IEEE Access, 2020(8): 228605-228618.
[3] 崔帅南, 彭宗举, 邹文辉, 等. 多特征融合的合成视点立体图像质量评价[J]. 电信科学, 2019, 35(5): 104-112.
CUI S N, PENG Z J, ZOU W H, et al. Quality assessment of synthetic viewpoint stereo image with multi-feature fusion[J]. Telecommunications Science, 2019, 35(5): 104-112.
[4] LI J, ZHANG H, WANG J, et al. Orientation-aware saliency guided JND model for robust image watermarking[J]. IEEE Access, 2019(7):41261-41272.
[5] XU Y F, ZHANG N, LI L, et al. Joint learning of super-resolution and perceptual image enhancement for single image[J]. IEEE Access, 2021(9):48446-48461.
[6] 邢亚芬, 殷海兵, 王鸿奎, 等. 基于视频时域感知特性的恰可察觉失真模型[J]. 电信科学, 2022, 38(2): 92-102.
XING Y F, YIN H B, WANG H K, et al. Video temporal perception characteristics based just noticeable difference model[J]. Telecommunications Science, 2022, 38(2): 92-102.
[7] BAE S H, KIM M. A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images[J]. IEEE Transactions on Image Processing, 2014, 23(8): 3227-3240.
[8] FRISTON K J, DAUNIZEAU J, KIEBEL S J. Reinforcement learning or active inference?[J]. PloS one, 2009, 4(7): e6421.
[9] FRISTON K. The free-energy principle: a unified brain theory?[J]. Nature Reviews Neuroscience, 2010, 11(2): 127-138.
[10] WU J J, SHI G M, LIN W S, et al. Just noticeable difference estimation for images with free-energy principle[J]. IEEE Transactions on Multimedia, 2013, 15(7): 1705-1710.
[11] WU J J, LI L D, DONG W S, et al. Enhanced just noticeable difference model for images with pattern complexity[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2682-2693.
[12] ZENG Z P, ZENG H Q, CHEN J, et al. Visual attention guide dpixel-wise just noticeable difference model[J]. IEEE Access, 2019(7): 132111-132119.
[13] LIU X Q, ZHAN X N, WANG M H. A novel edge-pattern-based just noticeable difference model for screen content images[C]//2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP). Piscataway: IEEE Press, 2020: 386-390.
[14] WANG H K, YU L, LIANG J H, et al. Hierarchical predictive coding-based JND estimation for image compression[J]. IEEE Transactions on Image Processing, 2020(30): 487-500.
[15] WATSON A B. DC Tune: A technique for visual optimization of DCT quantization matrices for individual images[C]//Sid International Symposium Digest of Technical Papers, Society for Information Display, 1993(24): 946-946.
[16] WEI Z Y, NGAN K N. Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(3): 337-346.
[17] ECKERT M P, BRADLEY A P. Perceptual quality metrics applied to still image compression[J]. Signal processing, 1998, 70(3): 177-200.
[18] ZHANG X H, LIN W S, XUE P. Improved estimation for just-noticeable visual distortion[J]. Signal Processing, 2005, 85(4): 795-808.
[19] KNILL D C, POUGET A. The Bayesian brain: the role of uncertainty in neural coding and computation[J]. TRENDS in Neurosciences, 2004, 27(12): 712-719.
[20] EASTTOM C. On the use of the SSIM algorithm for detecting intellectual property copying in web design[C]//2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC). Piscataway: IEEE Press: 2021: 0860-0864.
[21] LE C P, AUTRUSSEAU F. Subjective quality assessment IRCCyN/IVC database[EB]. 2005.
[22] PONOMARENKO N, JIN L, IEREMEIEV O, et al. Image database TID2013: peculiarities, results and perspectives[J]. Signal Processing: Image Communication, 2015, 30: 57-77.
[23] JUDD T, EHINGER K, DURAND F, et al. Learning to predict where humans look[C]//2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE Press, 2009: 2106-2113.
[24] WANG H K, YU L, YIN H B, et al. An improved DCT-based JND estimation model considering multiple masking effects[J]. Journal of Visual Communication and Image Representation, 2020(71): 102850.
[25] LI J L, YU L, WANG H K. Perceptual redundancy model for compression of screen content videos[J]. IET Image Processing, 2022, 16(6): 1724-1741.
[26] WANG J, LIU Y, WEI P, et al. Fractal image coding using SSIM[C]//2011 18th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2011: 241-244.

[作者简介]

骆琼华（1998– ），女，杭州电子科技大学通信工程学院硕士生，主要研究方向为视频感知编码。

王鸿奎（1990– ），男，博士，杭州电子科技大学通信工程学院讲师，主要研究方向为视觉感知理论、视频感知编码以及视频智能编码。

殷海兵（1974– ），男，博士，杭州电子科技大学通信工程学院教授，主要研究方向为数字视频编解码、图像和视频处理以及 VLSI 结构设计。

邢亚芬（1997– ），女，杭州电子科技大学通信工程学院硕士生，主要研究方向为视频感知编码。