

# 中文文本屏幕内容图像通用视频编码标准 编码感知失真研究

杨楷芳<sup>1,2,3</sup>, 晁学敏<sup>1</sup>, 蒙琴琴<sup>1</sup>, 公衍超<sup>4</sup>

(1. 陕西师范大学计算机科学学院, 710119, 西安; 2. 现代教学技术教育部重点实验室, 710062, 西安; 3. 陕西省教学信息技术工程实验室, 710119, 西安;  
4. 西安邮电大学通信与信息工程学院, 710121, 西安)

**摘要:** 为了探索最新一代通用视频编码标准(versatile video coding, VVC)对中文文本屏幕内容图像(text screen content image, TSCI)感知质量的影响,设计图像主观观测实验并基于VVC混合编码框架原理研究了中文TSCI的VVC编码感知失真。构建中文文本屏幕内容图像数据库(Chinese text screen content image dataset, CT-SCID),设计图像主观观测实验,分析VVC引起的中文TSCI感知失真类型及其发展路径;结合VVC的混合编码框架原理,理论分析并实验验证影响VVC编码的中文TSCI感知失真程度的因素;总结当前代表性的屏幕内容图像质量评价方法在中文TSCI VVC编码感知失真评测上的性能表现。实验结果表明:字体大小和对比度是影响中文TSCI VVC编码感知质量的重要因素,且中文TSCI的字体越小、对比度越低时,图像的感知质量等级越低;当前代表性的屏幕内容图像质量评价方法均无法给出完全符合人眼感知特性的质量评价结果。研究对于后续开发适用于中文TSCI的感知质量评价方法、高效编码方法等具有指导意义。

**关键词:** 中文文本屏幕内容图像;通用视频编码;感知失真;笔画

**中图分类号:** TN911.7 **文献标志码:** A

**DOI:** 10.7652/xjtuxb202404002 **文章编号:** 0253-987X(2024)04-0018-14

## Research on the Perceptual Distortion of Chinese Text Screen Content Image Using Versatile Video Coding

YANG Kaifang<sup>1,2,3</sup>, CHAO Xuemin<sup>1</sup>, MENG Qinqin<sup>1</sup>, GONG Yanchao<sup>4</sup>

(1. School of Computer Science, Shaanxi Normal University, Xi'an 710119, China; 2. Key Laboratory of Modern Teaching Technology, Ministry of Education, Xi'an 710062, China; 3. Shaanxi Provincial Teaching Information Technology Engineering Laboratory, Xi'an 710119, China; 4. School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

**Abstract:** To explore the influence of the state-of-the-art versatile video coding (VVC) on the perceptual quality of the Chinese text screen content image (TSCI), image subjective observation experiments were designed to study the perceptual distortion of Chinese TSCI using VVC based on the hybrid coding framework principles of VVC. A Chinese text screen content image dataset (CT-SCID) was constructed and image subjective observation experiments were designed to analyze the types and development paths of Chinese TSCI perceptual distortion caused by VVC.

收稿日期: 2023-05-29。 作者简介: 杨楷芳(1987—),女,副教授,硕士生导师;公衍超(通信作者),男,副教授,硕士生导师。 基金项目: 国家自然科学基金资助项目(62277036)。

网络出版时间: 2023-11-02

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20231102.1408.003>

By combining the hybrid coding framework principles of VVC, factors that affect the degree of perceptual distortion of Chinese TSCI using VVC were theoretically analyzed and experimentally verified. The performance of representative screen content image quality evaluation methods for evaluating the perceptual distortion of Chinese TSCI using VVC was summarized. Experimental results show that font size and contrast are crucial factors affecting the perceptual quality of Chinese TSCI using VVC. A smaller font size and lower contrast of Chinese TSCI shall lead to a lower perceptual quality of the image. The existing representative screen content image quality evaluation methods are unable to provide quality evaluation results that fully conform to human visual perception characteristics. The research has some guiding significance for the subsequent development of perceptual quality evaluation methods and efficient coding methods applicable to Chinese TSCI.

**Keywords:** Chinese text screen content image; versatile video coding; perceptual distortion; stroke

随着多媒体技术和云技术的快速发展,屏幕内容图像被广泛应用于远程会议、屏幕共享、在线教育等领域<sup>[1]</sup>。屏幕内容图像是由计算机或其他电子设备生成或渲染的图像,通常包含文字、图形、图表、图标、动画等内容<sup>[2]</sup>。在当前公开的屏幕内容图像数据库中,文字区域面积占比超过40%<sup>[1]</sup>。主要包含文字内容的屏幕内容图像又被称为文本类屏幕内容图像(text screen content image, TSCI)。相比于图形、图表、图标、自然场景等内容,文字作为与人类先验知识密切相关的符号,所携带的语义对图像信息的准确感知有重要影响。文献[1, 3]研究表明,人眼对文字内容的质量感知更加敏感。所以,TSCI的感知质量显著影响远程会议、在线教育等视频应用中用户的感知体验,是决定这些视频通信系统有效运行的关键因素之一。

在远程会议、在线教育等通信系统中,TSCI需要经过采集、编码、传输、存储、显示、分析等操作<sup>[4]</sup>。编码的目的是用更少的编码码率获得更高质量的重建TSCI。编码通常是有损的,这不可避免地在重建TSCI中引入失真,显著影响TSCI的感知质量。当前一些研究已经关注了TSCI的感知质量。文献[3]通过分析主观实验感知质量分数的统计特征发现文字区域的感知质量与图像总体感知质量的相关性更强,且文字的清晰度和完整性是人眼评价文本感知质量的重要依据。然后,采用锐度相似度和亮度相似度作为文字清晰度和完整性的测度,并提出有效的TSCI质量评价方法。文献[5]指出文字区域含有丰富的高对比度边界信息,这些信息与文本内容的准确表达密切相关,而小波变换可以提供多

尺度的边界信息。因此,文献[5]将文本图像从像素域变换到小波域,并基于小波系数的幅值、方差、熵等信息提出了有效的TSCI质量评价方法。与文献[5]类似,文献[2, 6-7]也关注到了文字边界信息对TSCI感知质量的重要影响,并分别提出了适用于编码失真的质量评价方法。不同的是,文献[2]提取的是文字边界对比度、边界宽度和边界方向特征,文献[6]提取的是文字边界的梯度和宽度特征,而文献[7]提取的是基于Gabor滤波器的时空Gabor特征张量模型(spatiotemporal Gabor feature tensor-based model, SGFTM)特征。文献[8]考虑屏幕内容图像中的文本形状规则且包含大量细线条和锐利笔画,因此提出利用像素的标准差分布特征衡量文本区域感知质量。面向由JPEG、JPEG2000标准压缩的图像,文献[9-11]分别提出了有效的质量评价方法。文献[9]采用梯度信息表征文字区域的结构特征,并进一步根据原始图像和失真图像的结构特征相似度得到文字区域的质量。文献[10]通过分析文本区域和图像区域标准差分布的差异,提出一种考虑局部标准差分布特征的质量评价方法。文献[11]则基于人眼对文本区域的先验知识,提出了考虑文本区域感知重要性的质量评价方法。但是当前研究还存在以下明显问题:

(1)都是面向的英文TSCI,未考虑中文TSCI。汉字是中华文化和信息传播的重要载体,是中文文本的符号系统。汉字作为一类由特定笔画构成的表意文字,也是目前世界上还在使用的人数最多的象形文字,其笔画特征是影响汉字语义信息表达的重要因素<sup>[12-13]</sup>。英文作为一类表音文字,其源于



拉丁字母,单个字母本身没有含义,字母的特定组合形成单词后才能表达语义<sup>[14]</sup>。因此,中英文文字系统的先天差别决定了中文 TSCI 和英文 TSCI 对应的编码失真特性也会存在明显不同,需要针对中文 TSCI 的编码感知特性进行针对性研究。

(2)都是面向早期的编码标准,未考虑最新的通用视频编码标准(versatile video coding, VVC)标准。现有研究<sup>[2, 3, 5-11]</sup>中已经关注的编码标准包括 1992 年发布的 JPEG 标准、2000 年发布的 JPEG2000 标准、2003 年发布的 H. 264/AVC 标准、2013 年发布的 HEVC 标准和 2017 年发布的 HEVC-SCC 标准。JPEG、JPEG2000、H. 264/AVC、HEVC 标准在设计时并未考虑屏幕内容图像的典型特性,所以针对屏幕内容图像的编码效率较低<sup>[15-16]</sup>。为了提升屏幕内容图像的编码效率,第一个面向屏幕内容图像/视频的编码标准,即 HEVC-SCC 标准,应运而生。近几年随着屏幕内容视频朝高清化、多维度等方向快速发展,HEVC-SCC 标准的编码效率已经很难满足实际需求。相应地,ITU-T 和 ISO/IEC 联合制定并发布了最新的通用视频编码(versatile video coding, VVC)标准<sup>[17]</sup>。相比于之前的标准,VVC 采用了大量先进的编码技术,例如改进的调色板模式、帧内块拷贝、块差分脉冲编码调制等。新的技术大幅提高了 VVC 编码屏幕内容视频的效率,同时也显著改变了屏幕内容图像的编码失真类型和感知表现。但是,当前针对 TSCI 的 VVC 编码感知失真研究还是空白。

本文首次聚焦中文 TSCI 的 VVC 编码感知失真研究,并首次构建中文文本屏幕内容图像数据库(Chinese text screen content image dataset, CT-SCID)。进一步地,结合图像主观观测实验和 VVC 混合编码框架原理,探索分析 VVC 引起的 TSCI 感知失真类型及其发展路径,理论分析及实验验证影

响感知失真程度的因素,并总结当前代表性的图像质量评价方法在评价这些感知失真时的性能表现。

## 1 中文文本屏幕内容图像数据库

### 1.1 数据库构建

目前,涉及 TSCI 编码感知质量评估的数据库包括 SIQAD<sup>[3]</sup>、SCID<sup>[2]</sup>、QACS<sup>[5]</sup>、SCVD<sup>[7]</sup>、CSCVQ<sup>[8]</sup>、SCD<sup>[18]</sup>。涉及的编码失真包括由 JPEG、JPEG2000、H. 264/AVC、HEVC 和 HEVC-SCC 标准产生的失真。这些数据库还存在以下明显问题:①包含的 TSCI 绝大多数为英文 TSCI,中文 TSCI 只有极少的几幅,且未对中文汉字的失真特性做针对性的分析总结;②未考虑 VVC 标准,数据库中没有包含 VVC 编码的失真图像。所以,为了便于后续针对中文 TSCI 开展 VVC 编码感知失真的研究,需要首先建立中文 TSCI 数据库。

在充分调研已有数据库指标及 TSCI 典型应用场景的基础上,本文构建了中文文本屏幕内容图像数据库(Chinese text screen content image dataset, CT-SCID),如图 1 所示。CT-SCID 共包括 55 幅原始图像,涉及远程会议、在线教育、屏幕共享、网页浏览等典型的 TSCI 应用场景,包含各类办公软件、学术文献、代码编辑、网页、新闻、广告、游戏等丰富内容。数据库还考虑了文本对比度、字体大小、字体类型等因素。文本对比度的范围为  $[139, 5\ 115]$ ,字体大小的范围为  $[327, 3\ 509]$ ,范围能够覆盖大部分实际场景。字体大小和文本对比度衡量测度的定义请分别参见第 3.1 节和第 3.2 节中相关内容。字体类型考虑了宋体、楷体、微软雅黑、黑体等主流字体。原始图像的空间分辨率为  $1\ 920 \times 1\ 080$ ,这是当前 TSCI 典型应用场景中硬件终端广泛支持的空间分辨率之一。图像颜色空间为 YCbCr 4:4:4,比特深度为 8。



图 1 CT-SCID 包含的原始图像

Fig. 1 Original images in the CT-SCID

图像作为由像素构成的二维矩阵,其空域细节信息的多少也是衡量图像特性的一个重要因素<sup>[19]</sup>。ITU-R BT. 1788 标准定义了空域感知信息(spatial perceptual information, SI)<sup>[19]</sup> 衡量图像的空域细节信息。为了适应 TSCI 纹理边缘较尖锐,甚至包含单像素边界的特点,本文计算 SI 时采用的是 Canny 算子,且计算的是二值图像的标准差。SI 越大表明图像包含的空域细节信息越多,图像越复杂。图 2 给出了 CT-SCID 数据库中 55 幅原始图像的 SI。可以看出,图中的数据点可以涵盖 SI 较大的范围,表明原始图像对应的空域细节信息分布较广泛,能够覆盖大部分实际场景。

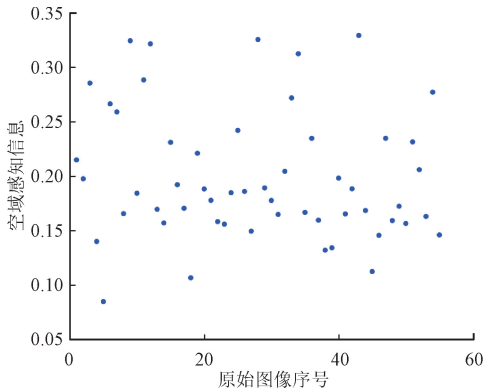


图 2 CT-SCID 中原始图像的 SI

Fig. 2 SI of the original images in CT-SCID

采用 VVC 官方推荐的测试模型 VTM16.2<sup>[20]</sup> 编码原始图像获得对应的失真图像。VVC 仍然沿用传统的基于预测变换量化熵编码的混合编码框架<sup>[17, 21]</sup>,但是在每个编码模块中都采用了更加先进的技术以提高视频编码的效率。例如,在预测模块中,VVC 采用了更灵活的编码单元块划分和更丰富的帧内帧间预测方向。在熵编码模块中,VVC 采用了更灵活更高效的上下文模型。但是,预测、变换、熵编码中的技术在原理上都是无损的,而量化才是引起视频编码失真的根本原因<sup>[4, 21]</sup>。一方面,VVC 采用多对一的量化映射机制,在减少变换系数取值空间的同时会显著降低视频细节信息的表达能力,导致视频清晰度降低,视频内容变模糊,引入编码失真。另一方面,考虑人类视觉系统对于视频细节信息感知不敏感,VVC 进一步采用了量化矩阵技术,对高频系数使用更大的量化步长(quantization step, QS),这易导致高频信息的丢失,从而使得视频内容进一步变模糊,引入编码失真。综上可知,相比于其他编码策略或技术,量化中的 QS 是影响编

码失真的最重要因素<sup>[21]</sup>。当前代表性的屏幕内容图像数据库<sup>[3, 5, 8, 18]</sup>也都是考虑调整 QS 来获得不同编码失真程度的图像。本文也是沿用这一思路。

在实际应用中,VTM 是通过选择量化参数(quantization parameter, QP)来控制 QS 的取值。VVC 规定的 QP 取值范围为  $-6(\eta-8) \sim 63$ ,其中  $\eta$  表示比特深度。韦伯-费希纳定理<sup>[22]</sup>表明,人眼不可能感知两幅图像间较小的质量差别,只有当质量差别超过一定阈值时人眼才能感知得到。当图像采用较小 QP 编码时,重建图像与原始图像之间的差别较小,人眼很难感知得到。所以,较小 QP 编码产生的重建图像感知质量都处于与原始图像感知质量一样的最好等级。当图像采用较大 QP 编码时,重建图像已经非常模糊,很难提取有用信息,这时不同 QP 对应的重建图像对应的感知质量等级是一样的,都处于最差等级。

基于以上人眼感知的先验知识可知,数据库中没有必要包含较小或较大 QP 对应的失真图像,因为它们的感知失真等级是确定的,且增加失真图像也会显著提高后续图像主观观测实验的成本。所以,本文考虑的 QP 范围为 32~60,采用 VTM16.2 编码原始图像获得每一个 QP 对应的失真图像。其他主要编码参数使用配置文件 encoder\_intra\_vtm.cfg 中的默认配置,档次为 main\_10\_444\_still\_picture。最终得到的 CT-SCID 共包含 1 595 幅失真图像。

## 1.2 失真图像主观观测实验

本节设计失真图像主观观测实验以获得失真图像的感知质量等级。主观观测实验涉及的关键因素和流程均严格按照 ITU-R BT. 500-13 标准<sup>[23]</sup>中的规定执行。具体地,实验共包括 30 名测试者,其中男性 17 名、女性 13 名,年龄在 20~40 岁之间。所有测试者都没有图像质量评价领域的研究经验,且经过检测均具有正常的视力。采用 side-by-side 方式,将原始图像和对应的失真图像同时显示给测试者,测试者依据平均意见分数(mean opinion score, MOS)五级量表给对应的失真图像质量评级。MOS 五级量表是 ITU-R BT. 500-13 标准推荐的且被广泛使用的图像感知质量评级测度,其将图像的感知质量划分为 5 个等级,并用 1~5 分表示,分数越高表示图像的感知质量越好。在实验开始前,会首先向测试者说明实验流程和注意事项。在实验过程中,当测试者持续观测时间达到 20 min 时,强制让测试者休息 5 min,以避免视觉疲劳。

通过以上过程可获得每一位测试者给每一幅失真图像打的感知质量分数。但是,受个体差异性等因素的影响,这些数据中可能会存在一些异常值。本文应用  $3\sigma$  准则<sup>[24]</sup>筛选并剔除异常值。 $n$  幅失真图像感知质量分数的均值  $\mu_n$  和标准差  $\sigma_n$  为

$$\mu_n = \frac{1}{\alpha} \sum_{m=1}^{\alpha} S_{m,n} \quad (1)$$

$$\sigma_n = \sqrt{\frac{1}{\alpha-1} \sum_{m=1}^{\alpha} (S_{m,n} - \mu_n)^2} \quad (2)$$

式中:  $S_{m,n}$  表示第  $m$  位测试者对第  $n$  幅失真图像打的感知质量分数;  $\alpha$  表示测试者数量。当某位测试者对应的感知质量分数处于  $(\mu_n - 3\sigma_n, \mu_n + 3\sigma_n)$  之外时,则认为其为异常点并将剔除。然后,计算剩余测试者对应的感知质量分数均值,此均值即为第  $n$  幅失真图像对应的 MOS。

图 3 为 CT-SCID 中所有失真图像对应的 MOS。可以看出,CT-SCID 数据库中失真图像的感知质量等级范围足够广,可用于后续中文 TSCI 的感知失真研究。

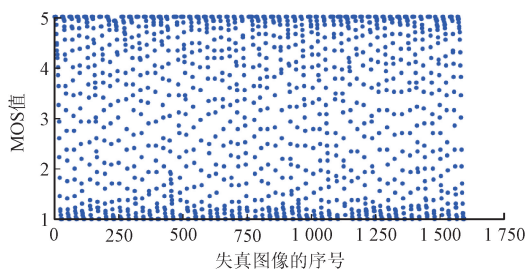


图 3 CT-SCID 中失真图像的 MOS 分布

Fig. 3 MOS distribution of distorted images in CT-SCID

## 2 中文 TSCI VVC 编码感知失真类型

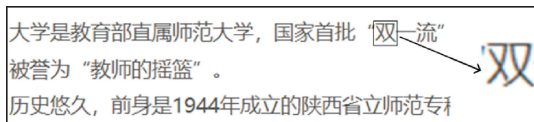
汉字作为由特定类型笔画构成的象形文字,其笔画信息对汉字语义的感知具有决定性作用。笔画信息属于图像的高频信息。包含 VVC 在内的传统基于混合编码框架的编码技术考虑人眼对自然图像高频信息不敏感,通常会使用压缩高频信息的方式实现数据压缩。但是,汉字笔画信息的变化却会对 TSCI 感知质量产生重要影响。本文研究发现,中文 TSCI 在经过 VVC 编码时,随着 QP 的增加,其对应重建图像中汉字的笔画会发生多种类型变化,对应产生多种类型的感知失真。本节将说明这些感知失真类型及其发展路径。

### 2.1 感知失真类型分析

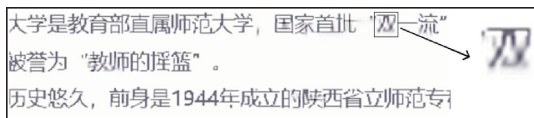
#### 2.1.1 笔画模糊

模糊是图像编码导致的一种常见失真。自然图

像和屏幕内容图像经过有损编码后都会出现模糊,但是具体的失真感知形式会有不同。图 4 以“WebPage02”图像中的区域为例,展示汉字对应的笔画模糊失真。图 4(a)、(b)分别为原始图像和 QP 为 49 时编码得到的失真图像,并将其中的“双”字放大展示。可以看出,原始图像中的“双”字笔画边界比较清晰,易识别其表示的语义,而失真图像中的“双”字笔画边界已经不分明,笔画相对模糊,出现了模糊失真。笔画模糊会增加汉字语义识别的难度。



(a) 原始图像



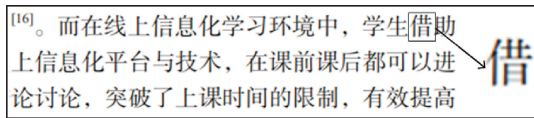
(b) QP 为 49 时对应的失真图像

图 4 “WebPage02”图像笔画模糊示例

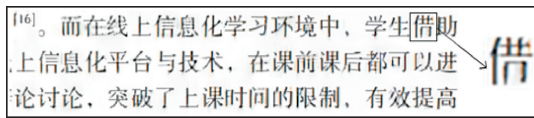
Fig. 4 Example of stroke blur on “WebPage02” image

#### 2.1.2 笔画丢失

除了笔画模糊外, TSCI 中的汉字还会出现笔画丢失失真。图 5 以“PDF01”图像中的区域为例,展示了汉字对应的笔画丢失失真。图 5(a)、(b)分别为原始图像和 QP 为 46 时编码得到的失真图像,并将其中的“借”字放大展示。易看出,原始图像中的“借”字笔画清晰且笔画结构完整,而失真图像中的“借”字虽然笔画较清晰,但第 9、10 笔顺对应的两个“横”缺失了。笔画丢失显著影响原汉字语义的准确感知,影响中文 TSCI 的感知质量。



(a) 原始图像



(b) QP 为 46 时对应的失真图像

图 5 “PDF01”图像笔画丢失示例

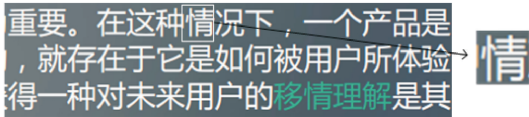
Fig. 5 Example of stroke lose on “PDF01” image

#### 2.1.3 笔画增加

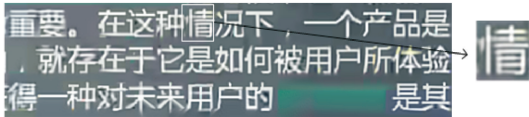
与笔画丢失对应, TSCI 中的汉字也会出现笔画增加失真。图 6 以“PPT01”图像中的区域为例,



展示了汉字对应的笔画增加失真。图 6(a)、(b)分别为原始图像和 QP 为 48 时编码得到的失真图像,并将其中的“情”字放大展示。易看出原始图像中的“情”字笔画清晰且笔画结构完整,而失真图像中的“情”字虽然笔画清晰,但其第 9 笔顺“横折钩”变为了“横折横”,导致“情”字的下半部分“月”变成了“目”。笔画增加也影响原汉字语义的准确感知,影响中文 TSCI 的感知质量。



(a)原始图像



(b)QP 为 48 时对应的失真图像

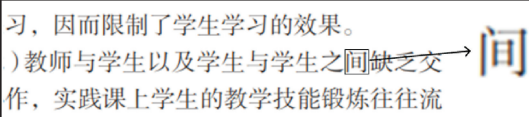
图 6 “PPT01”图像笔画增加示例

Fig. 6 Example of stroke adding on “PPT01” image

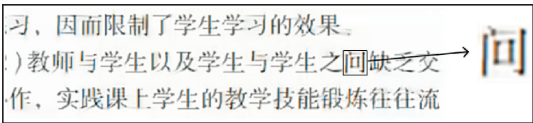
2.1.4 字符转换

前面提到的失真类型,即笔画模糊、笔画丢失、笔画增加,通常会影响汉字语义的识别难度,其导致的最严重情况是人眼难以识别汉字。但是,本节涉及的第 4 种失真,即字符转换,则会导致汉字语义的错误识别,对中文 TSCI 的感知质量产生严重影响。

当原始 TSCI 中的汉字经过编码后,汉字的笔画发生了变化,使得变化后的字符转换为了与原始汉字不同的其他汉字,本文将此种特殊失真类型定义为字符转换。图 7 以“PDF13”图像中的区域为例,展示了字符转换失真。图 7(a)、(b)分别为原始图像和 QP 为 45 时编码得到的失真图像,并将相应的汉字放大展示。原始图像中的“间”字笔画清晰且笔画结构完整,但其在失真图像中则转换为了“问”字,且“问”字的笔画也相对清晰。可以认为,字符转换是汉字笔画在增加或减少过程中出现的一种特殊情况,即增加或减少笔画后的符号与汉字字库中除原始汉字之外的某一个汉字的笔画结构趋近相同。区别于前 3 种失真类型,字符转换传递给人眼虚假语义信息,欺骗人眼对于汉字语义的感知,严重影响中文 TSCI 的感知质量。



(a)原始图像



(b)QP 为 45 时对应的失真图像

图 7 “PDF13”图像字符转换示例

Fig. 7 Example of character transformation on “PDF13” image

2.2 感知失真类型变化

原始中文 TSCI 在使用 VTM 编码时,随着 QP 增大,失真图像中汉字的笔画出现各种变化,从而产生前述的各种失真类型。为了便于分析这些失真类型的发展路径,将原始图像中的汉字定义为原始汉字字符(original Chinese character, OCC),将其对应应在失真图像中的汉字定义为失真汉字字符(distorted Chinese character, DCC)。表 1 直观地展示了中文 TSCI VVC 编码感知失真类型随 QP 增大的发展路径。图 8 给出了两个汉字形象的代表性示例。下面结合表 1、图 8 进行说明。

表 1 感知失真类型随 QP 增大的发展路径

| Table 1 Perceptual distortion type development path with increasing QP |         |      |           |
|------------------------------------------------------------------------|---------|------|-----------|
| 发展路径                                                                   | 失真类型    | 量化参数 | 示例        |
|                                                                        |         |      | 原始字符 失真字符 |
| 第 1 阶段                                                                 | 笔画感知无变化 | 32   | 生 生       |
|                                                                        | 笔画模糊    | 38   | 体 体       |
| 第 2 阶段                                                                 | 笔画丢失    | 41   | 片 片       |
|                                                                        | 笔画增加    | 47   | 有 有       |
|                                                                        | 字符转换    | 50   | 定 定       |
| 第 3 阶段                                                                 | 字符模糊    | 51   | 静 静       |
| 第 4 阶段                                                                 | 字符消失    | 63   | 据 据       |

第 1 阶段:笔画感知无变化。当 QP 较小时,图像高频信息损失较少。通过韦伯-费希纳定理可知,人眼很难感知到笔画信息的微小变化,主观上可认为 DCC 和 OCC 是一样的。

第 2 阶段:笔画模糊,笔画丢失,笔画增加,字符转换。当编码 QP 继续增大到一定程度时,其引起的笔画变化已经达到人眼感知的阈值。此时,DCC 对应产生笔画模糊、笔画丢失、笔画增加、字符转换



等失真。在本阶段存在两个重要感知现象:①定义第1个感知现象为空域耦合失真效应,即任一QP对应的某一个DCC有可能呈现出一种失真类型,也有可能同时呈现出多种失真类型;②定义第2个感

知现象为时域耦合失真效应,即随着QP增大,任一DCC可能会出现笔画模糊、笔画丢失、笔画增加、字符转换中的任意几个失真类型,而不是所有的失真类型都会出现。以上两个感知现象见图8。

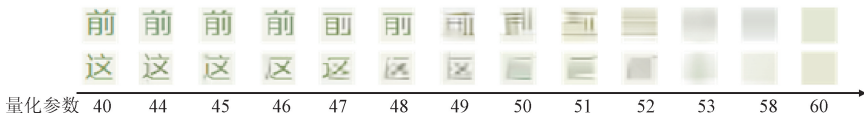


图8 “WebPage03”图像中的“前”和“这”感知失真类型发展路径

Fig. 8 Perceptual distortion types development path on “前” and “这” in “WebPage03” image

第3阶段:字符模糊。当编码QP继续增加时,图像高频信息损失严重,DCC的笔画信息严重模糊。

第4阶段:字符消失。当QP取值很大时,图像高频信息几乎全部损失,DCC对应的区域已经没有任何笔画信息。字符模糊和字符消失可以认为是笔画模糊的极端情况。

下面以图8为例直观地说明感知失真的变化路径。对于“WebPage03”图像中的“前”和“这”两个字,当QP较小(为40)时,DCC对应着第一个阶段,即笔画无变化阶段。随着QP增大,“前”和“这”对应的DCC先后进入第二个阶段。QP为46时,“这”对应的DCC同时出现了笔画模糊和字符转换两种失真,此即空域耦合失真效应。“前”对应的DCC先后出现了笔画丢失、笔画增加和笔画模糊,未出现字符转换失真,此即时域耦合失真效应。当QP继续增大到很大时,“前”和“这”对应的DCC先后进入字符模糊和字符消失阶段。

### 3 中文文本屏幕内容图像VVC编码感知失真影响因素

前一节已经说明了中文TSCI VVC编码失真类型及其随着QP增大的变化趋势。从信号处理的角度看,中文TSCI编码过程可以抽象为信源—技术—信宿的信号处理一般模型。中文TSCI是原始信源,其以像素域形式表达图像中的信息。编码器是信息处理技术,通过预测、变换、量化和熵编码等子技术达到数据压缩的目的。编码后的码流是信宿端接收和存储的数据,以压缩域形式表示图像中的信息。结合以上分析易得出以下结论:①在实际应用中,中文TSCI的编码感知失真通常在信宿端出现;②信宿端出现的编码感知失真肯定会同时受信源和编码技术的显著影响。VVC编码技术对TSCI

感知失真的影响主要体现为QP的影响,这部分在第1.1节和第2节已经分析。本节将从信源的角度分析影响中文TSCI VVC编码感知失真程度的重要因素。这些因素具体包括字体大小和文本对比度。

#### 3.1 字体大小

字体大小是汉字的一个重要属性。定义中文TSCI的字体大小 $P$ 为

$$P = \frac{1}{\beta} \sum_{c=1}^{\beta} (w_c h_c) \quad (3)$$

式中: $w_c$ 、 $h_c$ 分别表示图像中第 $c$ 个字符的宽度和高度; $\beta$ 表示图像中字符的数量。在计算 $P$ 之前采用投影分割法对图像中汉字进行分割<sup>[25]</sup>,从而得到每个字符。

图9(a)~(c)分别给出了 $P=578, 401, 245$ 时对应的3幅原始中文TSCI。对3幅图像使用VTM16.2编码,QP设置为49。得到的重建失真图像分别如图9(d)~(f)所示。原始图像及失真图像对应的SI数据也给出,以客观反映图像的纹理复杂度。由图9易知,字体大小显著影响中文TSCI编码重建图像的感知质量,且字体越小图像的感知质量等级越低。VVC标准采用基于预测、变换、量化和熵编码的混合编码框架。原始图像首先经过帧内预测技术去除图像中的空域冗余。帧内预测后的残差值是后续变换量化的输入。变换在去除一部分空域冗余的同时也使得能量更加集中。变换后的变换系数经过量化模块完成多对一的映射,显著减小了信息的取值区间。量化后的量化系数再经过熵编码,进一步去除熵冗余,最终达到数据压缩的目的。

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(a)原始图像, $P=578$ ,SI为0.363

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(b)原始图像,  $P=401$ , SI 为 0.403

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(c)原始图像,  $P=245$ , SI 为 0.422

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(d)失真图像, QP 为 49,  $P=578$ , SI 为 0.36

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(e)失真图像, QP 为 49,  $P=401$ , SI 为 0.394

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(f)失真图像, QP 为 49,  $P=245$ , SI 为 0.409

图 9 字体大小对于感知失真程度影响  
Fig. 9 Impact of font size on perceptual distortion degree

理论上,帧内预测技术是无损的,且单独使用该技术无法实现数据压缩,但是其对于提升后续变换、量化和熵编码等模块的效率至关重要。通常,图像纹理越复杂,临近像素间的相关性就越弱,帧内预测技术找到的最优预测值与当前编码像素之间的差别就越大,即预测残差就越大。如图 9(a)~(c)所示,图像中的字体越小,即单位面积上的笔画越密,图像的纹理就越复杂,其对应的 SI 也就越大。所以,字体越小的图像对应的预测残差值也就更大。图 10 给出了图 9(a)~(c) 3 幅原始图像在使用 VTM16.2、QP 为 49 编码时得到的帧内预测残差值分布。可以看出,字体越小的图像,预测残差值接近 0 的像素越少,即其对应的预测残差值相对偏大。进一步地,VVC 采用的离散余弦变换、离散正弦变换等变换技术,理论上是无损的,且满足能量守恒定律。所以,基于量化多对一的映射原理,在同等 QP 下,更大的预

测残差易导致更大的量化失真。VVC 采用的基于上下文的自适应二进制算术编码(context-based adaptive binary arithmetic coding,CABAC)等熵编码技术同样是无损的,所以更大的量化失真最终会导致重建图像中的失真更大,图像质量更差。

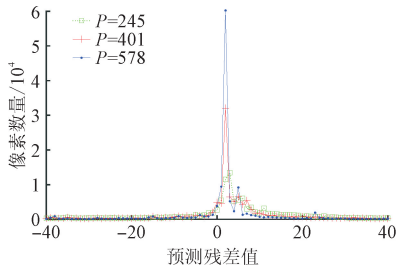


图 10 不同字体大小中文 TSCI 对应的帧内预测残差分布

Fig. 10 Distribution of intra prediction residuals for Chinese TSCI with different font sizes

当前,屏幕内容图像的质量评价是一个研究热点,一些高水平方法被相继提出。本文复现了一些代表性方法,包括 ESIM<sup>[2]</sup>、GFM<sup>[26]</sup>、GSIM<sup>[27]</sup>、GSS<sup>[28]</sup>、MDOGS<sup>[29]</sup>、MSEA<sup>[30]</sup>。图 11 为 12 幅测试图像,是由 VTM16.2 在 QP 为 49 时编码获得地, $P$  取值范围为 886~204。图 12 为不同的质量评价方法对 12 幅测试图像的测试结果。除了上述 6 种针对屏幕内容图像设计的质量评价方法,也给出了图像质量评价领域应用最广泛的两个传统质量评价测度 PSNR、SSIM<sup>[31]</sup> 的评价结果和 MOS 值。特别地,GSS 对应的分数越高说明图像质量越差,而其他方法对应的分数越高说明图像质量越好。

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况。

(a) $P=886$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况。受到课程时长的限制,往往会出现匆匆结束辩论的情况。

(b) $P=778$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况。

(c) $P=672$



传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况。教师根据班级情况将学生分成若干小组,再组织学生完成辩论。

(d)  $P = 634$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(e)  $P = 578$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(f)  $P = 501$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制。

(g)  $P = 469$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(h)  $P = 401$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(i)  $P = 325$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(j)  $P = 302$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(k)  $P = 245$

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

(l)  $P = 204$

图 11 QP 为 49 时不同字体大小中文 TSCI 的失真图像  
Fig. 11 Distorted images of Chinese TSCI with different font sizes when QP is 49

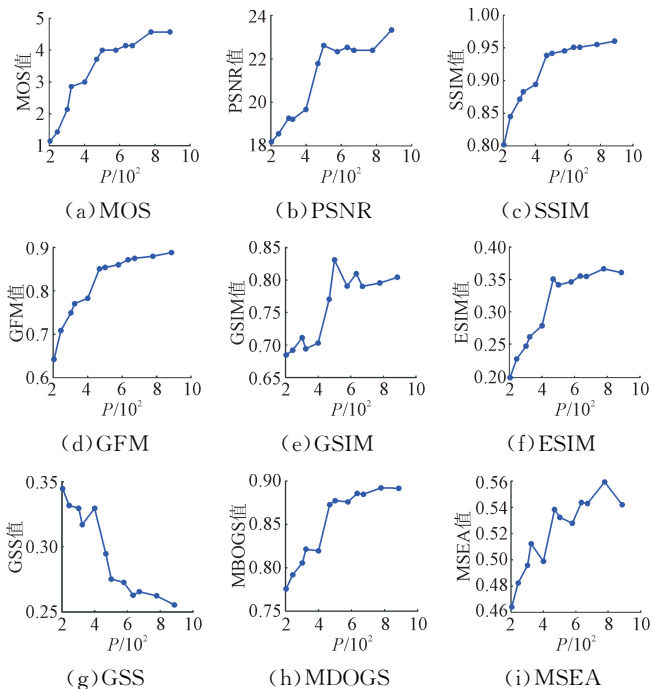


图 12 不同字体大小的中文 TSCI 失真图像质量评价结果  
Fig. 12 Quality assessment results of Chinese TSCI distorted images with different font sizes

对比图 11 和 12, 可以从宏观整体和微观细节两个角度得到以下结论: ①在同样 QP 下, 随着字体变小, 失真图像的质量在宏观整体上呈下降趋势, 图 12 复现的所有评价方法都能描述这一整体变化趋势; ②对于字体接近的一些图像, 其图像差别较少且不影响文本语义的感知, 基于韦伯-费希纳定理和文本感知的特点, 大部分测试者认为这些图像的感知质量是一样的。但是, 当前大部分质量评价方法很难准确描述这一微观差别。例如, 图 11(a)、(b) 两幅图像感知质量明显是一样的, 但图 12 中大部分测度认为这两幅图像的质量是不一样的, PNSN、GSS 和 MSEA 得出的分数差别甚至还较大。综上, 本文认为有必要针对中文 TSCI 的失真特性研究针对性的更加有效的质量评价方法, 且需要考虑字体大小的影响。



3.2 文本对比度

文本对比度是影响中文 TSCI 编码失真程度的另一个重要属性。应用中文 TSCI 的文本对比度  $D$  为

$$D = \frac{\sum_{i=1}^w \sum_{j=1}^h (\sum_{\epsilon=-1}^1 \sum_{\theta=-1}^1 (x_{i,j} - x_{i+\epsilon,j+\theta})^2)}{8\omega h - 6\omega - 6h + 4} \quad (4)$$

式中:  $x_{i,j}$  表示图像第  $i$  行第  $j$  列的像素值;  $w$ 、 $h$  分别表示图像的宽度和高度。

图 13(a)~(c) 分别展示了  $D=3\,542, 1\,704, 757$  时对应的 3 幅原始中文 TSCI。对这 3 幅图像分别使用 VTM16.2 编码, QP 为 46, 得到如图 13(d)~(f) 所示的重建失真图像。结合图像的 SI 数据可以看出, 在对比度变化影响下, SI 仍然可以有效衡量中文 TSCI 的纹理复杂度, 例如图 13(a)~(c) 包含的文字相同, 其对应的 SI 取值也一样。由图 13 易看出, 文本对比度显著影响中文 TSCI 编码重建图像的感知质量, 且对比度越小图像的感知质量等级越低。

1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单, 教师根据班级情况将学生分成若干小组, 再组织学生完成辩论。然而, 受到课程时长的限制, 往往会出现匆匆结束辩论的情况, 使得教学质量大打折扣。

在信息化教学环境下, 线上教学过程不受时间和地点的限制, 提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间, 特别是在这次疫情期间, 教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(a) 原始图像,  $D=3\,542$ , SI 为 0.379

1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单, 教师根据班级情况将学生分成若干小组, 再组织学生完成辩论。然而, 受到课程时长的限制, 往往会出现匆匆结束辩论的情况, 使得教学质量大打折扣。

在信息化教学环境下, 线上教学过程不受时间和地点的限制, 提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间, 特别是在这次疫情期间, 教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(b) 原始图像,  $D=1\,704$ , SI 为 0.379

1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单, 教师根据班级情况将学生分成若干小组, 再组织学生完成辩论。然而, 受到课程时长的限制, 往往会出现匆匆结束辩论的情况, 使得教学质量大打折扣。

在信息化教学环境下, 线上教学过程不受时间和地点的限制, 提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间, 特别是在这次疫情期间, 教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(c) 原始图像,  $D=757$ , SI 为 0.379

1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单, 教师根据班级情况将学生分成若干小组, 再组织学生完成辩论。然而, 受到课程时长的限制, 往往会出现匆匆结束辩论的情况, 使得教学质量大打折扣。

在信息化教学环境下, 线上教学过程不受时间和地点的限制, 提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间, 特别是在这次疫情期间, 教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(d) 失真图像, QP 为 46,  $D=3\,542$ , SI 为 0.375

1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单, 教师根据班级情况将学生分成若干小组, 再组织学生完成辩论。然而, 受到课程时长的限制, 往往会出现匆匆结束辩论的情况, 使得教学质量大打折扣。

在信息化教学环境下, 线上教学过程不受时间和地点的限制, 提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间, 特别是在这次疫情期间, 教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(e) 失真图像, QP 为 46,  $D=1\,704$ , SI 为 0.364

1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单, 教师根据班级情况将学生分成若干小组, 再组织学生完成辩论。然而, 受到课程时长的限制, 往往会出现匆匆结束辩论的情况, 使得教学质量大打折扣。

在信息化教学环境下, 线上教学过程不受时间和地点的限制, 提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间, 特别是在这次疫情期间, 教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(f) 失真图像, QP 为 46,  $D=757$ , SI 为 0.309

图 13 文本对比度对感知失真程度影响

Fig. 13 Impact of text contrast on the perceptual distortion degree

中文 TSCI 可以划分为前景汉字和背景两部分。并且, 相比于自然图像等其他内容, 汉字在内容特性上具有以下显著特点: 汉字笔画的亮度或颜色相对单一, 甚至经常出现一种亮度或颜色的情况。所以, 汉字笔画区域上的像素相关性较强。使用帧内预测编码汉字笔画上的像素时, 如果参考像素也来源于笔画上的其他像素, 则理论上其预测效率会较高, 对应的预测残差值较小。但是, 如图 13(a)~(c) 所示, 当图像文本对比度变小时, 文字笔画像素取值与背景像素取值越接近。背景像素对笔画像素最优帧内预测参考值选择的干扰增大, 笔画像素对应预测残差值变大的概率也会增大。图 14 给出了图 13(a)~(c) 3 幅原始图像在使用 VTM16.2、QP 为 46 编码时得到的帧内预测残差值分布。可以看出, 文本对比度越小的图像, 预测残差值接近 0 的像素越少, 即其对应的预测残差值相对偏大。如第 3.1 节分析, 在同等 QP 下, 更大的预测残差易导致更大

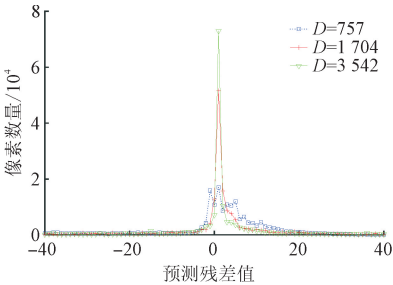


图 14 不同文本对比度中文 TSCI 对应的帧内预测残差分布

Fig. 14 Distribution of intra prediction residuals for Chinese TSCI with different text contrasts

的量化失真。更大的量化失真会导致重建图像中失真更大,图像质量更差。

图 15 为 12 幅测试图像,由 VTM16.2 在 QP 为 46 时编码获得, $D$  取值范围为 4 937~757,对应现实中常见的文本对比度。图 16 为不同的质量评价方法对应的测试图像的质量评价结果。

对比图 15 和 16 可以得到以下结论:①在同样 QP 下,随着文本对比度变小,失真图像的质量在宏观整体上呈下降趋势,而图 16 显示,除 PSNR 外的所有方法都能描述这一整体变化趋势。图 15(a)的感知质量显著好于图 15(l)的感知质量,但是图 15(a)的 PSNR 值反而比图 15(l)的 PSNR 值低,这显然是错误的评价结果。②对于文本对比度引起的失真,人眼通常更关注文本的清晰度是否影响文本语义的准确感知,而忽略背景对比度的变化<sup>[3]</sup>。当前的大部分质量评价方法很难准确描述对比度变化对中文 TSCI 感知质量的影响。例如,图 15(a)、(b)所示的两幅图像的差别很小,大部分测试者认为其感知质量是一样的,但是图 16 的所有质量评价方法得出的这两幅图像对应的质量分数显示两幅图像的质量均有差别,与人眼感知结果不符,甚至一些方法对应的质量分数差别较大,例如 GSIM、MSEA。综上所述可知,有必要针对中文 TSCI 的失真特性研究针对性的更加有效的质量评价方法,且需要考虑文本对比度的影响。

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(a) $D=4\ 937$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(b) $D=4\ 601$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(c) $D=4\ 449$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(d) $D=4\ 014$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(e) $D=3\ 542$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(f) $D=3\ 193$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(g) $D=2\ 768$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(h) $D=2\ 402$

#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(i) $D=2\ 064$

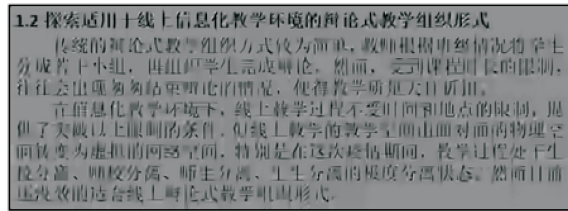
#### 1.2 探索适用于线上信息化教学环境的辩论式教学组织形式

传统的辩论式教学组织方式较为简单,教师根据班级情况将学生分成若干小组,再组织学生完成辩论。然而,受到课程时长的限制,往往会出现匆匆结束辩论的情况,使得教学质量大打折扣。

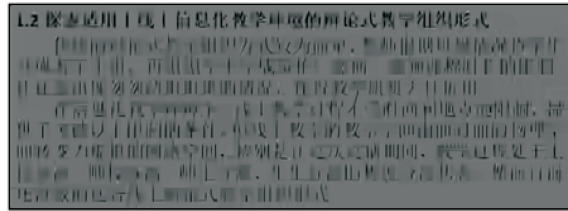
在信息化教学环境下,线上教学过程不受时间和地点的限制,提供了突破以上限制的条件。但线上教学的教学空间由面对面的物理空间转变为虚拟的网络空间,特别是在这次疫情期间,教学过程处于生校分离、师校分离、师生分离、生生分离的极度分离状态。然而目前还没效的适合线上辩论式教学组织形式。

(j) $D=1\ 704$





(k)D=1 229



(l)D=757

图 15 QP 为 46 时不同文本对比度中文 TSCI 的失真图像  
Fig. 15 Distorted images of Chinese TSCI with different text contrasts when QP is 46

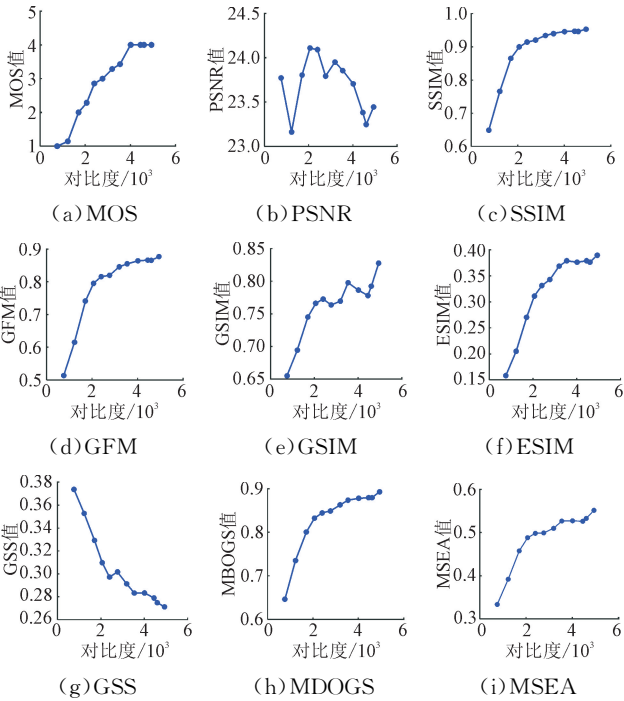


图 16 不同文本对比度的中文 TSCI 失真图像质量评价结果  
Fig. 16 Quality assessment results of Chinese TSCI distorted images with different text contrasts

4 结 论

本文聚焦中文 TSCT 图像的感知失真,基于构建的中文文本屏幕内容图像数据库,分析了中文 TSCI 的感知失真类型和发展路径。在此基础上,结合 VVC 混合编码架构原理对影响感知失真程度的因素进行了分析,对比了中文 TSCI 在经过 VVC

编码后,当前代表性的屏幕内容图像质量评价方法的失真性能表现,得到如下结论:

(1)作为由特殊笔画构成的象形文字,中文 TSCI VVC 编码感知失真可以从笔画变化的角度将失真类型进行有效分类,并分为笔画模糊、笔画丢失、笔画增加和字符变换 4 种类型;

(2)字符变换作为汉字编码时出现的一类特殊的失真类型,会传递错误的语义信息,欺骗人眼感知;

(3)随着 QP 由小到大变化,中文 TSCI VVC 编码感知失真类型的变化路径分为 4 个阶段,并在第二个阶段呈现空域和时域耦合失真效应;

(4)字体大小和文本对比度是影响中文 TSCI VVC 编码失真程度的关键因素,且字体越小或文本对比度越低,则对应失真图像的感知质量等级越低;

(5)当前提出的大部分质量评价方法,能够宏观整体地描述中文 TSCI VVC 编码失真图像的质量,但在微观细节上还存在提升空间,需要针对中文 TSCI VVC 编码失真图像研究更加有效的质量评价方法,且需要考虑字体大小与文本对比度的影响。

后续有价值的研究方向包括:适用于中文 TSCI VVC 编码失真图像的质量评价方法,优化中文 TSCI 感知率失真性能的 VVC 高效编码方法等。

参考文献:

[1] TANG Tong, LI Ling, WU Xiaoyu, et al. TSA-SCC: text semantic-aware screen content coding with ultra low bitrate [J]. IEEE Transactions on Image Processing, 2022, 31: 2463-2477.

[2] NI Zhangkai, MA Lin, ZENG Huanqiang, et al. ES-IM: edge similarity for screen content image quality assessment [J]. IEEE Transactions on Image Processing, 2017, 26(10): 4818-4831.

[3] YANG Huan, FANG Yuming, LIN Weisi. Perceptual quality assessment of screen content images [J]. IEEE Transactions on Image Processing, 2015, 24 (11): 4408-4421.

[4] GONG Yanchao, YANG Kaifang, LIU Ying, et al. Quantization parameter cascading for surveillance video coding considering all inter reference frames [J]. IEEE Transactions on Image Processing, 2021, 30: 5692-5707.

[5] WANG Shiqi, GU Ke, ZHANG Xiang, et al. Subjective and objective quality assessment of compressed screen content images [J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2016, 6 (4): 532-543.



- [6] YANG Jiachen, BIAN Zilin, ZHAO Yang, et al. Full-reference quality assessment for screen content images based on the concept of global-guidance and local-adjustment [J]. IEEE Transactions on Broadcasting, 2021, 67(3): 696-709.
- [7] CHENG Shan, ZENG Huanqiang, CHEN Jing, et al. Screen content video quality assessment: subjective and objective study [J]. IEEE Transactions on Image Processing, 2020, 29: 8636-8651.
- [8] LI Teng, MIN Xionghuo, ZHAO Heng, et al. Subjective and objective quality assessment of compressed screen content videos [J]. IEEE Transactions on Broadcasting, 2021, 67(2): 438-449.
- [9] FANG Yuming, YAN Jiebin, LIU Jiaying, et al. Objective quality assessment of screen content images by uncertainty weighting [J]. IEEE Transactions on Image Processing, 2017, 26(4): 2016-2027.
- [10] JIANG Xuhao, SHEN Liquan, FENG Guorui, et al. An optimized CNN-based quality assessment model for screen content image [J]. Signal Processing: Image Communication, 2021, 94: 116181.
- [11] JAKHETIYA V, GU Ke, LIN Weisi, et al. A prediction backed model for quality assessment of screen content and 3-D synthesized images [J]. IEEE Transactions on Industrial Informatics, 2018, 14 (2): 652-660.
- [12] 王存睿, 丁阳, 刘宇, 等. 融合笔画语义和注意力机制的汉字字体生成算法 [J]. 计算机辅助设计与图形学学报, 2022, 34(8): 1229-1237.  
WANG Cunrui, DING Yang, LIU Yu, et al. Chinese font generation from stroke semantic and attention mechanism [J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(8): 1229-1237.
- [13] 张积家. 汉字词认知过程中整体与部分关系论 [J]. 自然辩证法通讯, 2002, 24(3): 91-94.  
ZHANG Jijia. On the relation of entirety and components in identifying process of Chinese characters and words [J]. Journal of Dialectics of Nature, 2002, 24 (3): 91-94.
- [14] 苏培成. 现代汉字学的学科建设 [J]. 语言文字应用, 2007(2): 2-11.  
SU Peicheng. The discipline building of modern Chinese character study [J]. Applied Linguistics, 2007 (2): 2-11.
- [15] 李维, 任鹏, 赵凡, 等. 高效视频编码中帧内码率控制改进算法 [J]. 西安交通大学学报, 2019, 53(4): 79-84.  
LI Wei, REN Peng, ZHAO Fan, et al. A modified method for intra-coding rate control in high efficiency video coding [J]. Journal of Xi'an Jiaotong University, 2019, 53(4): 79-84.
- [16] 宋蓓蓓, 何帆, 马穗娜, 等. 一种高精度的自适应码率控制图像压缩算法 [J]. 西安交通大学学报, 2022, 56(2): 198-206.  
SONG Beibei, HE Fan, MA Suina, et al. An image compression algorithm with high-precision adaptive rate control [J]. Journal of Xi'an Jiaotong University, 2022, 56(2): 198-206.
- [17] BROSS B, WANG Yekui, YE Yan, et al. Overview of the versatile video coding (VVC) standard and its applications [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31 (10): 3736-3764.
- [18] SHI Sheng, ZHANG Xiang, WANG Shiqi, et al. Study on subjective quality assessment of Screen Content Images [C]//2015 Picture Coding Symposium (PCS). Piscataway, NJ, USA: IEEE, 2015: 75-79.
- [19] ITU. Methodology for the subjective assessment of video quality in multimedia applications; BT. 1788 [S]. Geneva: ITU, 2007: 1-13.
- [20] ITU ISO/IEC. VTM reference software for VVC [EB/OL]. [2022-05-03]. [https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware\\_VTM/-/tree/VTM-16.2](https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-16.2).
- [21] BROSS B, CHEN Jianle, OHM J R, et al. Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC) [J]. Proceedings of the IEEE, 2021, 109(9): 1463-1493.
- [22] 郭嘉骏, 姜求平, 邵枫. 分形理论引导的图像临界差异感知阈值估计 [J]. 中国图象图形学报, 2022, 27 (11): 3303-3315.  
GUO Jiajun, JIANG Qiuping, SHAO Feng. Fractal-guided JND threshold estimation for natural images [J]. Journal of Image and Graphics, 2022, 27(11): 3303-3315.
- [23] ITU. Methodology for the subjective assessment of the quality of television pictures; BT. 500-13 [S]. Geneva: ITU, 2012: 1-46.
- [24] 王信峰. 概率论与数理统计 [M]. 北京: 清华大学出版社, 2016.
- [25] 魏传义, 陈勤, 张旻. 基于投影的文本图像版面分割算法研究 [J]. 现代计算机, 2016(10): 33-38.  
WEI Chuanyi, CHEN Qin, ZHANG Min. Research on document image layout segmentation algorithm based on projection [J]. Modern Computer, 2016 (10): 33-38.

[26] NI Zhangkai, ZENG Huanqiang, MA Lin, et al. A Gabor feature-based quality assessment model for the screen content images [J]. IEEE Transactions on Image Processing, 2018, 27(9): 4516-4528.

[27] LIU Anmin, LIN Weisi, NARWARIA M. Image quality assessment based on gradient similarity [J]. IEEE Transactions on Image Processing, 2012, 21(4): 1500-1512.

[28] NI Zhangkai, MA Lin, ZENG Huanqiang, et al. Gradient direction for screen content image quality assessment [J]. IEEE Signal Processing Letters, 2016, 23(10): 1394-1398.

[29] FU Ying, ZENG Huanqiang, MA Lin, et al. Screen content image quality assessment using multi-scale difference of Gaussian [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(9): 2428-2432.

[30] YANG Qi, MA Zhan, XU Yiling, et al. Modeling the screen content image quality via multiscale edge attention similarity [J]. IEEE Transactions on Broadcasting, 2020, 66(2): 310-321.

[31] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

(编辑 陶晴 武红江)