

面向机器识别-人类感知的联合振动触觉编码

房颖, 徐艺文, 赵铁松

(福州大学福建省媒体信息智能处理与无线传输重点实验室, 福建 福州 350001)

摘要: 为了精确地传输信号内容含义, 实现智能识别与信号重建, 针对振动触觉信号, 提出了一种面向机器识别-人类感知的联合编码方案。在编码端, 将三维振动信号转化为一维信号, 采用短时傅里叶变换提取信号的语义信息, 并实现语义信息高效压缩与表征。在解码端, 基于语义信息采用全卷积神经网络实现触觉的智能识别; 同时, 将原始信号与基于语义信息重构信号的残差值作为语义信息的补偿, 逐步提高重构信号的质量, 满足人类感知需求。实验结果表明, 所提方案用较低比特率的语义信息实现触觉识别, 同时在满足人类感知需求情况下, 触觉数据的压缩效率有所提高。

关键词: 触觉; 语义信息; 感知质量; 联合编码; 智能识别; 振动触觉

中图分类号: TP37

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023099

Joint vibrotactile coding for machine recognition and human perception

FANG Ying, XU Yiwen, ZHAO Tiesong

Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information, Fuzhou University, Fuzhou 350001, China

Abstract: In order to accurately transmit the content meaning of vibrotactile signals and achieve intelligent recognition and signal reconstruction, a joint vibrotactile coding scheme for machine recognition and human perception was proposed. At the encoding end, the original three-dimensional vibrotactile signals were converted into one-dimensional signals. Then the semantic information of the signals was extracted using a short-time Fourier transform before being effectively compressed and transmitted. At the decoding end, a fully convolutional neural network was used to intelligently recognize based on the semantic information. The difference between the original signals and the reconstructed signals based on semantic information was used as compensation for the semantic information, and the quality of the reconstructed signals was gradually improved to meet human perceptual needs. The experimental results show that the proposed scheme achieve tactile recognition with semantic information at a lower bit rate while improving the compression efficiency of tactile data, thus satisfying human perceptual needs.

Keywords: haptic, semantic information, perceptual quality, joint coding, intelligent recognition, vibrotactile

0 引言

6G技术、人工智能(AI, artificial intelligence)技术、物联网技术以及传感器技术的融合发展将推

动多维度沉浸式交互以及智能信息网络的发展。传统以视听为核心的多媒体应用已逐渐不能满足现有用户的体验需求以及机器分析任务需求, 亟须引入新的感官信息。触觉作为人类环境感知的重要交

收稿日期: 2023-01-18; 修回日期: 2023-04-18

通信作者: 赵铁松, t.zhao@fzu.edu.cn

基金项目: 国家自然科学基金资助项目(No.62171134); 福建省自然科学基金资助项目(No.2022J02015)

Foundation Items: The National Natural Science Foundation of China (No.62171134), The Natural Science Foundation of Fujian Province (No.2022J02015)

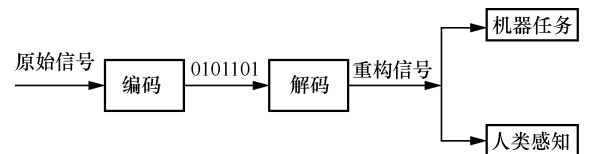
互方式, 能够传递硬度、大小、纹理、形状、温度、压力等视觉和听觉无法传递的信息, 极大地细化用户服务感受, 提高用户感知体验, 并为智能识别任务提供有力的辅助信息, 快速区分视觉上相似的物体, 提高智能认知程度^[1]。因此, 触觉技术在远程医疗、线上购物、在线教育、远程工业控制等领域具有良好的应用前景。

触觉包括位于关节和韧带内的动觉感知和位于皮肤真皮层及表皮层的肤感。动觉感知包括力、扭矩、位置、速度、角速度等信息。肤感包括静态压力、振动、摩擦、表面纹理、皮肤拉伸、三维形状、热觉和痛觉等信息^[2]。其中, 与材料表面粗糙度有关的振动触觉信号在现有的触觉应用引起了工业界和学术界的关注。为了满足真实的触觉感知需求, 通常需要使用较高的采样率来采集振动触觉信号, 并且随着触觉应用所需自由度数量的增加, 信号数据量将成比例增加^[3]。因此, 有必要研究高效的振动信号数据压缩方案。

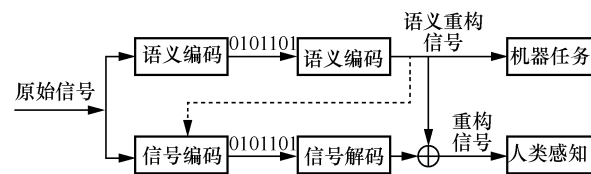
目前, 对于振动信号数据的压缩主要包括感知编码、变换编码和混合感知编码, 这三类编解码器在减少信号失真方面具有较好的表现^[4]。其研究重点是如何在有限的比特率下获得更高的触觉感知质量, 以兼顾用户体验与通信质量, 主要是为人类触觉感知服务的压缩方法。随着通信与 AI 紧密融合, 智能体之间要完成更多智能分析任务, 如对目标物体的分类识别等任务, 这就需要有效提取信号的语义信息并实现高效的压缩以提高语义传输的有效性^[5]。因此如何在对信号语义信息进行有效提取与编解码的同时满足机器与人类的需求成为数据压缩的一大挑战。

传统上, 为了降低数据存储及传输的成本, 通常会将输入信号进行编码, 得到压缩数据, 然后将压缩数据进行解码, 得到与原始信号相似的重构信号以满足人类感知或机器任务的需求, 如图 1(a)所示。在这种方法中, 机器的智能识别必须在信号重构后才能完成^[6], 对识别的实时性具有一定的影响, 而且重构信号的失真可能导致识别任务的失败。为了能够同时实现智能识别与信号重构, 兼顾机器任务与人类感知需求, 一些研究^[7-8]采用可伸缩的数据编码算法, 如图 1(b)所示, 通过提取信号的语义特征, 对语义信息进行编码, 产生的特征位流作为基本层数据用于机器任务, 同时使用原始信号与语义信号的差值信号作为增强层数据实现满足人类感

知的信号重构。这种方式可以使用低比特率数据提供及时的智能分析, 以满足机器任务的需要, 在此基础上通过增加额外的比特流可以获得与原始信号相似的重构信号, 满足人类感知的需求。Choi 等^[9]提出了一个分层的端到端可扩展的图像编解码器, 其中基本层用于目标检测, 增强层结合基本层用于数据重构, 结果表明可扩展编解码器在保证机器视觉任务性能及重构图像信号质量的情况下可以节省 37%~80% 的码率。Huang 等^[10]提出了人机友好的视频压缩方案, 通过学习语义表示 (LSR, learned semantic representation) 的方法来提取时间上相邻帧之间的语义信息用于动作识别, 再利用增强层中原始信号与 LSR 的残差值进行信号重构, 可以很好地满足人类感知和机器分析的需要。Xia 等^[11]提出了一种可扩展视频编解码器, 利用关键节点作为语义信息进行人体动作的识别, 并基于预测和生成模型来进行信号的重构, 以更少的比特率获得更好的重建视频质量和更高的动作识别准确率。



(a) 传统面向机器任务与人类感知的编码方案



(b) 面向机器任务与人类感知的联合编码方案

图 1 面向机器任务与人类感知的编码方案

目前, 已有针对视频、图像语义信息的提取、压缩以及语义与感知的联合编码的研究, 但还没有针对触觉信号的相关研究。在实际触觉应用中, 触觉语义信息的有效提取可作为触觉控制指令进行传输进而被机器所理解, 这有助于触觉任务中低时延、高可靠性的性能需求, 同时以语义信息为基础进行触觉信号的重构有利于在保证信号质量的同时减少数据量。因此, 基于语义信息与人类感知的触觉信号的联合编码研究有其必要性。

本文以振动触觉信号为基础, 基于 IEEE P1918.1.1 触觉编解码标准化组织提供的振动触觉信号标准数据集, 提出了一种面向机器识别-人类感知的联合振动触觉编码方案。根据振动触觉信号

特点提取其语义信息并采用相应编码方法进行语义信息的压缩,然后采用深度学习算法,基于语义信息实现智能识别任务;同时获得原始信号与语义重构信号的差值信号,作为语义信息的补充,实现满足人类体验质量的信号重构。具体来说,本文贡献主要总结为以下三点。

1) 实现振动触觉信号语义信息的提取和有效的压缩。

2) 提出基于语义信息表征与人类感知的触觉信号的联合编码方案及其框架。语义信息作为基础层信号,满足人类感知的补偿信号作为增强层信号,两者处于不同的比特流中,可以同时满足智能识别与人类感知的需求。

3) 基于标准数据集,采用基于全卷积神经网络的时间序列分类算法,使用较低比特率的语义信息数据较好地实现了触觉信号的智能分类,并且在保证满足人类感知需求的情况下,提高触觉数据的压缩效率。

1 相关研究

当触摸物体表面时,物体表面会产生细微的振动,即振动触觉信号,这些信号携带大量关于表面材料特性的信息,构成了触觉感知的一个重要信息来源。振动触觉信号通常可以采用激光多普勒测温仪、高带宽摩擦仪或加速度传感器来获取。一般情况下,振动触觉信号通常发生在开环通信场景中,对时延具有一定的容忍性。目前,对于振动信号的编码算法主要包括感知编码、变换编码以及混合感知编码。

感知编码利用触觉感知理论和心理学来消除感知冗余。一种方法是基于韦伯定理的死区编码技术,当信号的变化范围超出可感知的阈值时进行信号的传输,否则信号将不被传输。Liu 等^[12]设计了一个数据驱动的振动信号的压缩函数,采用非均匀量化降低信号比特率。另一种方法则是利用振动触觉编码的频率掩蔽效应来减少数据量,Chaudhari 等^[13]基于此提出了一种比特率可扩展的触觉编解码器,实现数据压缩的同时保证信号没有明显的失真。

变换编码利用离散余弦变换或离散小波变换将振动信号从时间域转换到频率域,通过量化去除振幅低于阈值的信号,减少原始信号中的关联性,再进行熵编码,压缩信号数据量。文献^[14]将粗糙度建模为

一个高度剖面,并使用离散余弦变换编码实现了 4:1 的压缩率,同时保证信号具有可接受的感知质量。

混合感知编码将感知编码与变换编码相结合,通过保留可感知信号信息,来降低所需的数据传输速率。基于稀疏线性预测的感知振动触觉信号压缩(PVC-SLP, perceptual vibrotactile-signal compression based on sparse linear prediction)^[15]算法则采用了加速敏感度函数以构建触觉敏感度模型,并在线性预测方案中对残差和预测系数都引入了稀疏性约束,在频率域中用触觉灵敏度函数进行量化从而达到数据压缩的目的。基于感知小波量化的振动编解码器(VC-PWQ, vibrotactile signal compression based on perceptual wavelet quantization)^[16]利用感知的绝对阈值和掩蔽现象,动态提取信号的必要属性,同时使用离散小波变换有效地去除输入信号的相关性,最后结合嵌入式均匀量化器、多级树集合分裂算法及算术编码,得到了较好的触觉编码性能。

目前,对于振动信号数据的压缩主要针对人类触觉感知,其目的是更好地保证用户的沉浸式体验,但缺少面向智能任务的语义信息提取与数据压缩。本文提出面向机器识别与人类感知的联合振动触觉编码算法,使用低比特率语义信息数据实现智能分类任务,在此基础上增加更多比特率进一步重建信号,以满足人类感知的需要。

2 联合编码方案

IEEE P1918.1.1 触觉编解码标准化组织提供了 9 类不同表面材料在不同运动速度下的 280 个振动信号轨迹的标准触觉数据库^[17]。这些振动数据采用加速度表征,加速度计连接在手指上或集成在工具上。为了模拟真实世界的振动触觉感应,振动触觉信号被分离成 3 个连续的信号,每个信号包括 X、Y 和 Z 轴上的振动信号。振动信号采样率保持在 2 800 Hz,数据样本的振幅范围为 0~1.5,为了表示振动触觉强度轻微的变化,每个样本都以 16 bit 表示。

根据振动信号的特性,本文提出面向机器识别与人类感知的联合编码方案。本节主要描述了联合编码方案的框架,介绍了方案核心模块的实现方法。

2.1 联合编码方案框架

为了能够实现语义信息的提取与表征,并对其进行高效的压缩,兼顾机器任务与人类感知需求,本文提出了面向机器识别-人类感知的联合振动触觉编码方案,如图 2 所示。

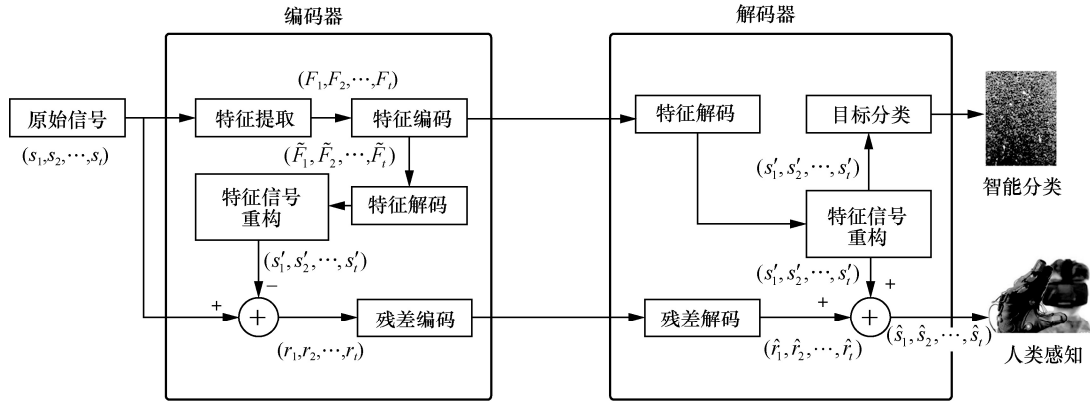


图2 面向机器识别-人类感知的联合振动触觉编码方案

首先，根据 IEEE P1918.1.1 触觉编解码标准化组织建议，采用 DFT321 技术^[18]将三维的振动信号 $\{x_t, y_t, z_t\}$ 转换为一维信号 s_t 。由于机器感受器对高频振动的方向基本不敏感，将三维高频振动信号转换为一维信号时人类不会出现明显的感知失真，并且不影响信号的频谱功率。

其次，在编码端，对一维信号进行特征提取获得语义信息 F_t ，再对语义信息进行量化、熵编码，形成低比特率的基本层语义信息比特流，用于特征信号的重构以支持机器任务的完成。过程如式(1)~式(3)所示。

$$(F_1, F_2, \dots, F_t) = f(s_1, s_2, \dots, s_t) \quad (1)$$

$$Z_t = Q_1(F_t) \quad (2)$$

$$F_t' = E_1(Z_t) \quad (3)$$

其中， $f(\cdot)$ 表示对原始信号的特征提取函数， $Q_1(\cdot)$ 表示对语义信息的量化函数， $E_1(\cdot)$ 表示对语义信息量化值的熵编码。为了满足人类感知需求，需要在语义信息之外增加额外的比特流对其进行补偿，因此对编码的语义信息采用 $D_1(\cdot)$ 进行熵解码、 $Q_1^{-1}(\cdot)$ 函数进行反量化后，重构信号 s_t' ，如式(4)~式(6)所示。将一维振动信号与重构的特征信号的残差值 r_t 进行量化、熵编码，形成增强层信号比特流，如式(7)~式(9)所示。

$$Z_t = D_1(F_t') \quad (4)$$

$$\tilde{F}_t = Q_1^{-1}(Z_t) \quad (5)$$

$$(s_1', s_2', \dots, s_t') = f^{-1}(\tilde{F}_1, \tilde{F}_2, \dots, \tilde{F}_t) \quad (6)$$

$$r_t = s_t - s_t' \quad (7)$$

$$R_t = Q_2(r_t) \quad (8)$$

$$r_t' = E_2(R_t) \quad (9)$$

其中， $Q_2(\cdot)$ 为对残差信号的量化函数， $E_2(\cdot)$ 为残差信号量化值的熵编码。编码端传送的比特流为语义信息码流 F_t' 和残差信号码流 r_t' 。

最后，在解码端，根据式(4)~式(6)对基本层语义信息比特流进行熵解码、反量化，以及基于语义信息的信号重构，再采用目标分类网络模型进行触觉信号的智能分类任务；同时，对增强层残差信号进行熵解码、反量化，结合特征信号 $(s_1', s_2', \dots, s_t')$ 与残差信号 \hat{r}_t ，重构高质量的触觉信号 \hat{s}_t ，以满足人类沉浸式多媒体体验，如式(10)~式(12)所示。

$$R_t = D_2(r_t') \quad (10)$$

$$\hat{r}_t = Q_2^{-1}(R_t) \quad (11)$$

$$\hat{s}_t = s_t' + \hat{r}_t \quad (12)$$

基于该框架，可使用较少比特的语义信息完成面向机器的智能识别任务，同时使用残差信号的补偿获得满足人类需求的高质量触觉信号。

2.2 特征信号的提取与重构

触觉信号的特征提取可使用递归神经网络提取信号的时间语义特征或者可采用类似音频信号的处理方法对触觉信号进行时频分析^[19]。文献[20-21]基于时频分析进行特征提取，完成触觉信号的分类任务，因此本文方案将触觉信号做短时傅里叶变换，分析其时频特点，并提取特征值。

虽然振动触觉信号在本质上是非稳态的，但在实际的信号建模中，可以假设其具有短期的稳态性。因此，振动触觉信号可被窗口化，分成不重叠的帧，确保每一帧内的统计特性基本不变。根据数据分析，振动触觉信号的标准数据库采样

率为 2 800 Hz，当帧长为 200 个样本（约 71 ms）时，每一帧中记录的加速度基本保持静止，因此本文方案采用汉明窗对触觉信号进行帧化，设置汉明窗的长度为 200 个样本点，窗位为 100 个样本点，对每一帧做快速傅里叶变换，频域信号在时间上堆叠获得时频谱图，如图 3 所示。从图 3 可明显观察到，该信号的频率成分主要集中于 600 Hz 以下，其保留了振动触觉信号的大部分能量。因此，本文方案保留 600 Hz 以下的低频数据作为振动触觉信号特征。

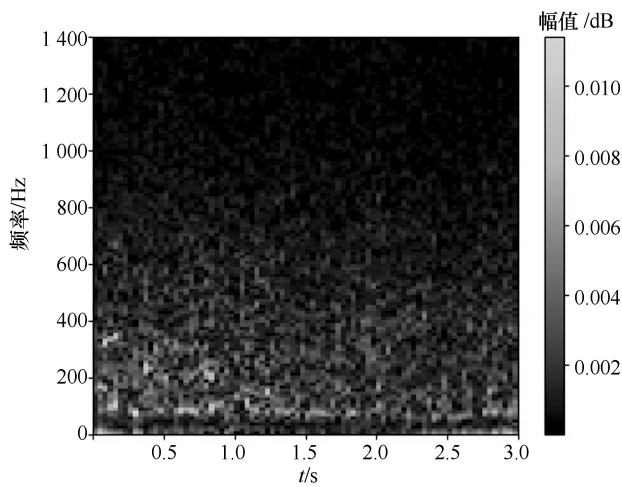


图 3 振动触觉信号的短时傅里叶变换的时频谱图

解码端利用振动触觉信号的语义信息进行重构时，对于未传输的高频特征补 0，然后利用短时傅里叶反变换获得其重构信号。

2.3 特征信号的编码

触觉信号经过短时傅里叶变换后，可以获得其低频信号的振幅与相位特征，为了进一步减少特征信息的比特率，本文方案对其进行量化及熵编码。

在振动触觉感知中，感知死区阈值随着信号振幅的增加而增加，因此，本文方案对语义信息进行非均匀量化处理。出于输出比特率和重构质量之间的均衡考虑，本文方案将低频信号的振幅和相位数据的精度设为 12 位，然后采用压扩量化编码，将每个样本数据精度设为 7 位，最后对其进行哈夫曼编码。

首先，将低频信号的振幅与相位数据采用均匀量化，数据精度设为 12 位。为了进一步调整数据输出比特率，引入了量化参数 Q ，用以调整均匀量化间隔，进而实现对数据压缩率的调整。量化间隔 L 的设置如式(13)所示。

$$L = \frac{Q(x_{\max} - x_{\min})}{2^{12}} \quad (13)$$

其中， x_{\max} 与 x_{\min} 分别表示振幅或相位的最大值与最小值。在本文方案中，信号振幅的取值范围为 $(0, 1)$ ，相位的取值范围为 $(-\pi, \pi)$ 。随着量化参数 Q 的增加，量化误差会有所增加，但数据比特数会减少。

然后，采用非均匀量化方案，将每个样本用 7 位编码表示。第 1 位表示样本值的正负极性，后面的 6 位分为段落码和段内码两部分，用于表示量化值的绝对值。其中，第 2 位~第 4 位是段落码，表示 8 种非均匀划分的段落；其他 3 位为段内码，表示每一段落内的 8 个均匀划分的间隔。段落码和段内码的编码规则分别如表 1 和表 2 所示。

表 1 段落码编码规则

段落码	段落范围
111	(2 047, 4 096)
110	(1 024, 2 047)
101	(512, 1 023)
100	(256, 511)
011	(128, 255)
010	(64, 127)
001	(32, 63)
000	(0, 31)

表 2 段内码编码规则

段内码	量化间隔
111	7
110	6
101	5
100	4
011	3
010	2
001	1
000	0

本文方案首先判断样本值 x 的正负极性，然后用式(14)获得 P 值，确定样本值 x 的绝对值所处的段落范围，根据段落码编码规则确定样本值 x 的段落码，最后用式(15)得到 p 值，获知样本值 x 的绝对值的段内区间，根据段内码编码规则确定样本值所处的段内码。由此，获得样本值 x 的 7 位编码表示。由于振幅信号都为正，因此不需要增加极性码位，只需采用 6 位编码表示，相位采用 7 位编码表示。

$$P = \frac{|x|}{L} \quad (14)$$

$$p = \left\lfloor \frac{P - S_{\min}(P)}{S_{\max}(P) - S_{\min}(P)} \times 2^3 \right\rfloor \quad (15)$$

其中, $S_{\max}(P)$ 与 $S_{\min}(P)$ 分别表示样本值所处段落范围的最大值与最小值。

最后, 用熵编码器对压括编码后的数据进行无损编码, 进一步降低数据流的比特率, 同时保证信号的质量。本文方案中振幅的哈夫曼编码如表 3 所示, 相位的哈夫曼编码如表 4 所示。表 4 中, 0 与 64 采用相同的哈夫曼编码是因为 0 与 64 都表示量化后的 0 值。在编码过程中, 对于重复的码值采用行程编码, 因此, 本文方案引入 NULL 标识, 用来表示当前时刻数据与前一时刻数据相同。NULL 标识之后使用一个 5 位二进制码表示后续重复码连续出现的个数 M 。当 M 小于 33 时, 使用 (NULL, BIN($M-1$)) 来表示重复出现的码值, 其中, BIN(\cdot) 表示数值的二进制; 当 M 大于 32 时, 则使用 (NULL, BIN(\cdot)) 来表示重复出现的码值, 如 (NULL, BIN(31)), ..., (NULL, BIN($M-31N$)), 其中, $N = \left\lfloor \frac{M}{31} \right\rfloor$, $\lfloor \cdot \rfloor$ 表示向下取整。利用行程编码可进一步压缩数据量。

表 3 振幅的哈夫曼编码

量化值	哈夫曼编码
NULL	1
0	01
1	001
2	0001
3	00001
4	000001
⋮	⋮

表 4 相位的哈夫曼编码

量化值	哈夫曼编码
NULL	1
0	01
1	001
2	00000100
3	000000101
⋮	⋮
64	01
⋮	⋮

2.4 残差信号编码

基于提取的语义信息, 可以使用低比特率数据提供及时的分析结果, 满足机器任务的需要, 但为了满足人类感知的需求, 需要在语义信号的基础上增加额外的比特流以丰富信号细节, 生成与原始信号相似的重构信号。本文方案采用原始振动触觉信号与基于语义信息重构的信号之间的差值作为语义信息的补偿数据, 以提高重构信号质量。

由于触感信号为一维时间序列信号, 其残差值在时序上依然存在着相关性, 因此本文基于预测编码来实现残差信号的编解码。基于门控循环单元 (GRU, gate recurrent unit) 的残差信号预测编码方案如图 4 所示。

在编码端, 利用 $t-1$ 和 $t-2$ 时刻的残差信号的重构值 \hat{r}_{t-2} 和 \hat{r}_{t-1} , 并根据式(16)预测 t 时刻的残差信号 \tilde{r}_t , 由式(17)获得预测值 \tilde{r}_t 与实际值 r_t 的预测误差值 Δr_t , 再对 Δr_t 进行非均匀量化及熵编码, 获得残差信号的输出码流。在解码端, 对预测误差值 Δr_t 的码流进行熵解码及反量化得到 $\Delta r_t'$, 利用式(16)获得预测值 \tilde{r}_t , 根据式(18)重构残差信号。

$$\tilde{r}_t = l(\hat{r}_{t-2}, \hat{r}_{t-1}) \quad (16)$$

$$\Delta r_t = r_t - \tilde{r}_t \quad (17)$$

$$\hat{r}_t = \tilde{r}_t + \Delta r_t' \quad (18)$$

预测编码中预测器的性能极大影响编码性能。由于 GRU 预测模型具有非线性特征学习的优势, 被证明在时序信号 (如振动触觉信号) 中有很高的预测性能^[22], 因此本文方案采用 GRU 递归神经网络模型作为残差信号的预测器。如图 4 所示, 该网络模型包括 2 个 GRU 及一个全连接层 Dense。其中, 2 个 GRU 的输入分别为 $t-2$ 和 $t-1$ 时刻的残差信号的重构值 \hat{r}_{t-2} 和 \hat{r}_{t-1} 。每个 GRU 包括重置门 Gate R 和更新门 Gate Z, 用于获取时间序列的依赖关系。 h_1 和 h_2 分别表示上一时刻的隐藏状态, h_3 表示当前时刻的隐藏状态, w_{r-i} 、 w_{z-i} 以及 w_{h-i} ($i=1,2$) 分别表示第 i 个 GRU 中重置门、更新门以及上一时刻隐藏状态的权重矩阵。

预测误差采用非均匀量化方案, 与语义信息量化方案相同 (见 2.3 节), 本文方案预测误差变化范围为 $(-0.5, 0.5)$ 。

预测误差量化后, 结合哈夫曼编码与行程编码将预测误差进一步压缩, 其哈夫曼编码如表 5 所示, 详细方法见 2.3 节熵编码部分。

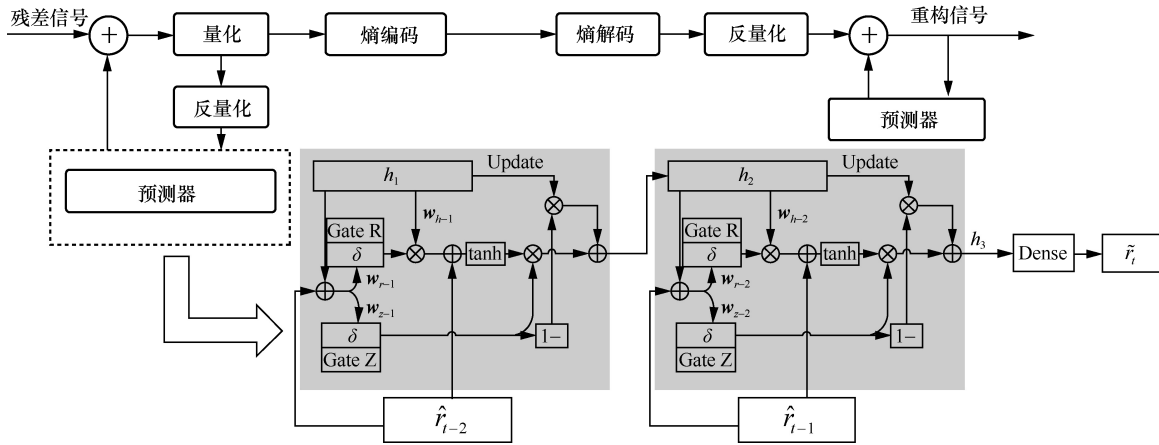


图 4 基于 GRU 的残差信号预测编码方案

表 5 残差信号的预测误差的哈夫曼编码

量化值	哈夫曼编码
NULL	1
0	01
1	001
2	0001
3	000000110
⋮	⋮
64	01
⋮	⋮

2.5 分类算法

全卷积神经网络（FCN）被证明在端到端方式下可以显著提高语义分割的计算效率和预测性能，在时间序列的分类任务中有较好的表现^[23]，因此本文采用 FCN 来实现触觉信号的分类任务。标准数据库提供了 9 类不同表面材料使用不同工具在不同运动速度下的振动信号轨迹。本文方案设定 FCN 的输入为语义信息的重构信号，其保留了信号最主要的特征信息，同时去除了语义冗余，更有利于智

能分类任务的完成。FCN 的输出标签为 9 类不同表面材料。本文采用的 FCN 模型如图 5 所示。

FCN 由 3 个基本卷积块组成，每个基本卷积块包括一个卷积层 Conv2D 和一个批量归一化（BN, batch normalization）层，采用修正线性单元（ReLU, rectified linear unit）激活函数。其中，卷积层用于提取信号特征信息，BN 用于防止梯度消失或爆炸，加快训练速度并帮助提高泛化能力。3 个卷积层分别使用大小为 8、5、3 的一维卷积核，3 个基本卷积块中的滤波器大小分别为 128、256、128。输入信号在经过 3 个基本卷积块后进行全局平均池化，以减少权重数量，最后结果由 softmax 层生成。

3 实验与性能分析

本文基于 IEEE P1918.1.1 触觉编解码标准化组织提供的振动触觉信号标准数据库从数据重构及智能分类两方面分析了联合编码性能。标准数据库中，实验人员使用不同工具在 9 类不同表面材料上以不同速度运动，形成 280 个振动信号轨迹，每条轨迹包括 3 360~8 400 个样本，每个样本以 16 位精度表示。

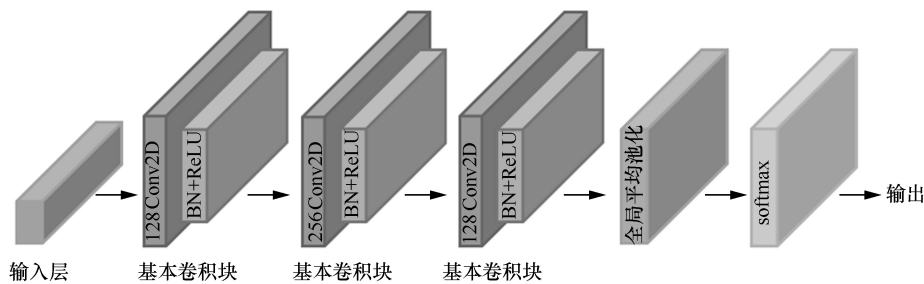


图 5 FCN 模型

3.1 数据重构性能

为了定量地衡量编解码器的性能，本文采用压缩率 (CR, compression rate) 衡量压缩性能，信噪比 (SNR, signal noise ratio) 和峰值信噪比 (PSNR, peak signal to noise ratio) 衡量信号重构质量，时频相似性 (ST-SIM, spectral temporal similarity) 从感知频谱和时间特性两方面评估原始信号与重构信号的相似性，衡量基于感知的信号重构质量^[24]，因其与主观质量评价结果高度相关，成为触感数据编码性能的重要指标之一。CR、SNR、PSNR 及 ST-SIM 的计算式分别如式(19)~式(22)所示。

$$CR = \frac{\text{原始信比特率}}{\text{压缩信比特率}} \quad (19)$$

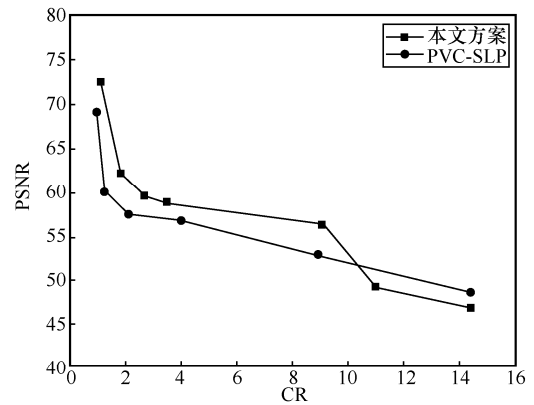
$$SNR = 10 \lg \left(\frac{\|x\|^2}{\|\tilde{x} - x\|^2} \right) \quad (20)$$

$$PSNR = 10 \lg \left(\frac{A^2}{\|\tilde{x} - x\|^2} \right) \quad (21)$$

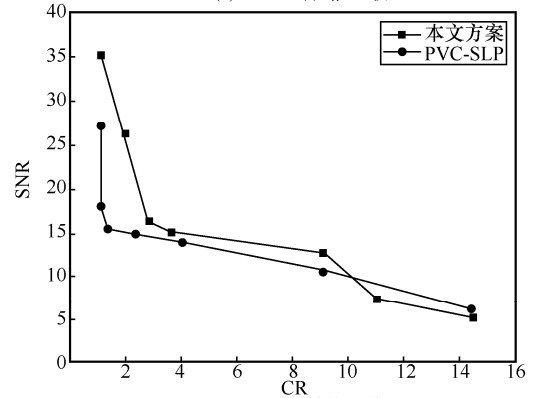
$$ST-SIM = \left(\frac{1}{N_s} \sum_{k=1}^{N_s} S-SIM(k) \right)^{1-\eta} \left(\frac{1}{N_s} \sum_{k=1}^{N_s} T-SIM(k) \right)^\eta \quad (22)$$

其中， x 为原始信号， \tilde{x} 为重构信号， A 为振动触觉信号峰值，设为 6.0^[15]， N_s 为 S-SIM 和 T-SIM 分段数。因为人类振动触觉主要与时间特征有关，系数 η 设为 $\frac{2}{3}$ 。

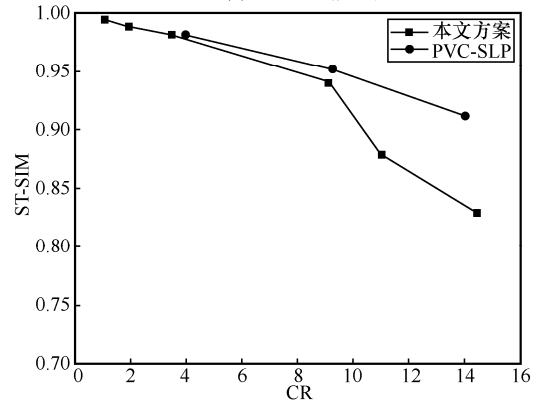
本文方案通过调整语义信息中信号的振幅、相位以及残差信号三者的量化参数 Q 实现量化精度的调整，进而改变压缩比特率，但同时也会对重构信号的质量也必然产生影响。PVC-SLP 目前被 IEEE P1918.1.1 触觉编解码标准化组织列为触觉编码标准，因此本节将本文方案的信号重构性能与 PVC-SLP 的信号重构性能进行对比，其结果如图 6 所示，本文方案在 CR 较小的情况下，PSNR、SNR 以及 ST-SIM 优于 PVC-SLP，但在 CR 增大的情况下其信号失真大于 PVC-SLP。这是由于 PVC-SLP 只针对人类感知质量进行优化，因此可以在大压缩率的情况下保证信号重构质量。而联合编码同时考虑了机器任务与人类感知质量两方面的目标优化，两者互相影响，为了保证机器任务的准确性，压缩性能就会有所下降。



(a) PSNR的性能比较



(b) SNR的性能比较



(c) ST-SIM的性能比较

图 6 本文方案与 PVC-SLP 的信号重构性能比较

3.2 数据分类性能

为了衡量联合编码的分类性能，本文采用准确率作为分类性能评价标准。

本文方案通过调整语义信息的振幅及相位的量化参数，获得语义信息数据集，将数据集以 7:3 的比例划分为 FCN 模型训练集和测试集。为了消除奇异样本数据导致的不良影响，本文使用数据集的平均值和标准偏差对训练集和测试集进行归一化处理。在模型训练过程中，损失函数为分类交叉熵，采用自适应矩估计 (Adam, adaptive moment estimation) 作为优化器，学习率初始值

设置为 0.01，动态调整学习率，其中学习率缩放系数设置为 0.5，当迭代 50 次模型性能没有改变时，学习率自动进行调整，学习率最小值设置为 0.001。由图 7 及图 8 可知，训练集的分类准确率可达 95.09%，测试集的分类准确率可达 85.90%。基于该模型，本文测试了语义信号在不同压缩比 CR 下分类准确率。首先，调整量化参数 Q 获得不同压缩比 CR；然后，基于不同压缩比下的语义信号进行重构，基于 FCN 实现分类，结果如图 9 所示。从图 9 可知，随着压缩比 CR 增大，分类准确率下降，但在数据量明显小于原始信号的情况下仍然保持较好的分类性能。本文方案所提取的语义信息保留了信号的内容特征，而对语义信息的压缩方案也在尽量保留了主要语义信息的同时减少了需要传输的比特数。同时，本文方案与 CNN 模型^[25]进行了对比，由图 9 可知，本文方案的分类准确率优于 CNN 模型。

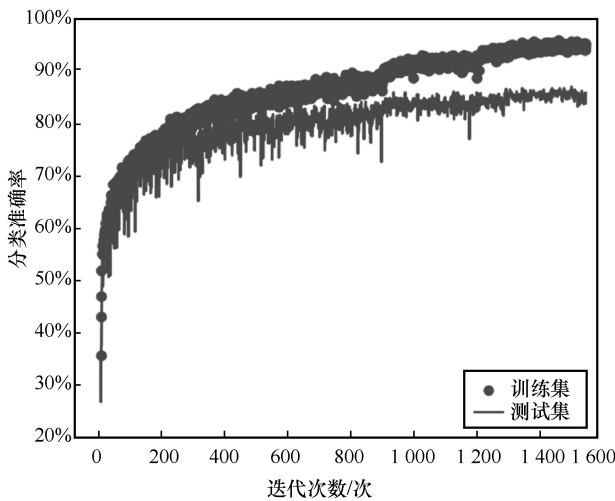


图 7 训练集和测试集的分类准确率

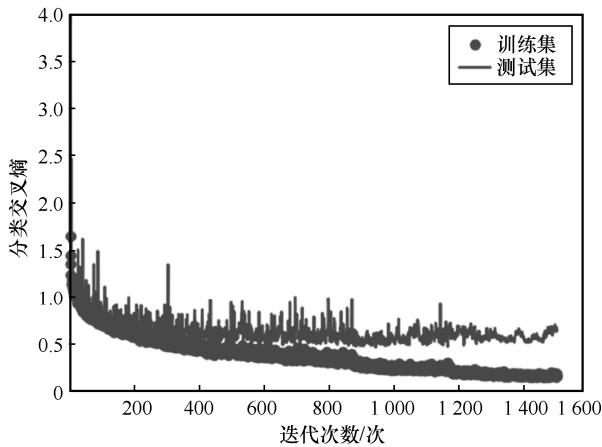


图 8 训练集和测试集的分类损失

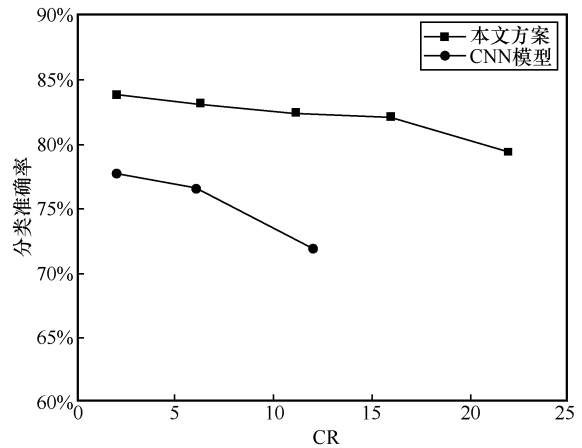


图 9 本文方案与 CNN 模型分类准确率比较

4 结束语

人工智能技术、物联网技术以及传感器技术的发展推动了触觉信息在未来的广泛应用。触觉信号在有效提高用户的多媒体沉浸式体验的同时能够提高机器智能分析任务的准确性。但随着触觉信号数据量的增加，需要开发高效的编解码算法。本文以振动触觉信号为基础，提出了一种面向机器识别-人类感知的联合振动触觉编码方案，根据振动触觉信号特点，利用短时傅里叶变换提取其语义信息，对其进行高效的压缩编码，减少了用于机器识别的信号冗余，使用低比特率语义信息数据较好地实现了智能分类任务，并在语义信息基础上增加比特率补偿进一步重建信号，以满足人类感知的需要。

参考文献:

- [1] HUANG H, HSIEH M H. Tactile emotional coding: the perceptual linking of vibrotactile stimuli with basic emotions[C]//Proceedings of 2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII). Piscataway: IEEE Press, 2020: 134-137.
- [2] ANTONAKOGLU K, XU X, STEINBACH E, et al. Toward haptic communications over the 5G tactile Internet[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 3034-3059.
- [3] XU X, CIZMECI B, SCHUWERK C, et al. Haptic data reduction for time-delayed teleoperation using the time domain passivity approach[C]//Proceedings of 2015 IEEE World Haptics Conference (WHC). Piscataway: IEEE Press, 2015: 512-518.
- [4] STEINBACH E, STRESE M, EID M, et al. Haptic codecs for the tactile Internet[J]. Proceedings of the IEEE, 2019, 107(2): 447-470.
- [5] 刘传宏, 郭彩丽, 杨洋, 等. 面向智能任务的语义通信: 理论、技术和挑战[J]. 通信学报, 2022, 43(6): 41-57.
- LIU C H, GUO C L, YANG Y, et al. Intelligent task-oriented semantic communications: theory, technology and challenges[J]. Journal on

- Communications, 2022, 43(6): 41-57.
- [6] DUAN L Y, LIU J Y, YANG W H, et al. Video coding for machines: a paradigm of collaborative compression and intelligent analytics[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2020, 29: 8680-8695.
- [7] HU Y Y, YANG S, YANG W H, et al. Towards coding for human and machine vision: a scalable image coding approach[C]//Proceedings of 2020 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2020: 1-6.
- [8] LIU K, LIU D, LI L, et al. Semantics-to-signal scalable image compression with learned reversible representations[J]. International Journal of Computer Vision, 2021, 129(9): 2605-2621.
- [9] CHOI H, BAJIĆ I V. Scalable image coding for humans and machines[J]. IEEE Transactions on Image Processing, 2022, 31: 2739-2754.
- [10] HUANG Z M, JIA C M, WANG S S, et al. HMFVC: a human-machine friendly video compression scheme[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, doi: 10.1109/TCSVT.2022.3207596.
- [11] XIA S F, LIANG K, YANG W H, et al. An emerging coding paradigm vcm: a scalable coding approach beyond feature and signal[C]//Proceedings of 2020 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2020: 1-6.
- [12] LIU X, DOHLER M. A data-driven approach to vibrotactile data compression[C]//Proceedings of 2019 IEEE International Workshop on Signal Processing Systems (SiPS). Piscataway: IEEE Press, 2020: 341-346.
- [13] CHAUDHARI R, SCHUWERK C, DANAEI M, et al. Perceptual and bitrate-scalable coding of haptic surface texture signals[J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 9(3): 462-473.
- [14] OKAMOTO S, YAMADA Y. Perceptual properties of vibrotactile material texture: effects of amplitude changes and stimuli beneath detection thresholds[C]//Proceedings of 2010 IEEE/SICE International Symposium on System Integration. Piscataway: IEEE Press, 2011: 384-389.
- [15] HASSEN R, GÜLECYÜZ B, STEINBACH E. PVC-SLP: perceptual vibrotactile-signal compression based-on sparse linear prediction[J]. IEEE Transactions on Multimedia, 2021, 23: 4455-4468.
- [16] NOLL A, NOCKENBERG L, GÜLECYÜZ B, et al. VC-PWQ: vibrotactile signal compression based on perceptual wavelet quantization[C]//Proceedings of 2021 IEEE World Haptics Conference (WHC). Piscataway: IEEE Press, 2021: 427-432.
- [17] KIRSCH J, NOLL A, STRESE M, et al. A low-cost acquisition, display, and evaluation setup for tactile codec development[C]//Proceedings of 2018 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE). Piscataway: IEEE Press, 2018: 1-6.
- [18] LANDIN N, ROMANO J M, MCMAHAN W, et al. Dimensional reduction of high-frequency accelerations for haptic rendering[C]//International Conference on Human Haptic Sensing and Touch Enabled Computer Applications. Berlin: Springer, 2010: 79-86.
- [19] 李昂, 陈建新, 魏昕, 等. 面向 6G 的跨模态信号重建技术[J]. 通信学报, 2022, 43(6): 28-40.
- LI ANG, CHEN J X, WEI X, et al. 6G-oriented cross-modal signal reconstruction technology[J]. Journal on Communications, 2022, 43(6): 28-40.
- [20] SAJEEVAN N, ARATHI N M, ARAVIND S R, et al. Surface material classification using acceleration signal[C]//Proceedings of International Conference on Communication and Computational Technologies. Berlin: Springer, 2021: 49-58.
- [21] ZHENG H T, FANG L, JI M Q, et al. Deep learning for surface material classification using haptic and visual information[J]. IEEE Transactions on Multimedia, 2016, 18(12): 2407-2416.
- [22] XIAO J Q, ZHOU Z Y. Research progress of RNN language model[C]//Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). Piscataway: IEEE Press, 2020: 1285-1288.
- [23] WANG Z G, YAN W Z, OATES T. Time series classification from scratch with deep neural networks: a strong baseline[C]//Proceedings of 2017 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2017: 1578-1585.
- [24] HASSEN R, STEINBACH E. Subjective evaluation of the spectral temporal SIMilarity (ST-SIM) measure for vibrotactile quality assessment[J]. IEEE Transactions on Haptics, 2020, 13(1): 25-31.
- [25] JI M Q, FANG L, ZHENG H T, et al. Preprocessing-free surface material classification using convolutional neural networks pretrained by sparse autoencoder[C]//Proceedings of 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP). Piscataway: IEEE Press, 2015: 1-6.

[作者简介]



房颖 (1981-), 女, 福建福州人, 博士, 福州大学讲师, 主要研究方向为多媒体通信、计算机触觉。



徐艺文 (1976-), 男, 福建漳州人, 博士, 福州大学教授、硕士生导师, 主要研究方向为媒体信息智能处理与传输。



赵铁松 (1984-), 男, 河北衡水人, 博士, 福州大学教授、博士生导师, 主要研究方向为多媒体通信、计算机视觉、计算机触觉等。