



## 基于视-触跨模态感知的智能导盲系统

朱文霖, 刘华平, 王博文, 孙富春

引用本文:

朱文霖, 刘华平, 王博文, 等. 基于视-触跨模态感知的智能导盲系统[J]. 智能系统学报, 2020, 15(1): 33–40.

ZHU Wenlin, LIU Huaping, WANG Bowen, et al. An intelligent blind guidance system based on visual–touch cross–modal perception[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(1): 33–40.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201908015>

## 您可能感兴趣的其他文章

### 基于生成对抗网络的机载遥感图像超分辨率重建

Super-resolution reconstruction of airborne remote sensing images based on the generative adversarial networks  
智能系统学报. 2020, 15(1): 74–83 <https://dx.doi.org/10.11992/tis.202002002>

### 视听觉跨模态表面材质检索

Audiovisual cross-modal retrieval for surface material  
智能系统学报. 2019, 14(3): 423–429 <https://dx.doi.org/10.11992/tis.201804030>

### 触觉手势情感识别的超限学习方法

Extreme learning machine for emotion recognition of tactile gestures  
智能系统学报. 2019, 14(1): 127–133 <https://dx.doi.org/10.11992/tis.201804029>

### 基于卷积神经网络的盲文音乐识别研究

Research on braille music recognition based on convolutional neural networks  
智能系统学报. 2019, 14(1): 186–193 <https://dx.doi.org/10.11992/tis.201805002>

### PG-RNN:一种基于递归神经网络的密码猜测模型

PG-RNN: a password-guessing model based on recurrent neural networks  
智能系统学报. 2018, 13(6): 889–896 <https://dx.doi.org/10.11992/tis.201712006>

### 利用多模态U形网络的CT图像前列腺分割

Prostate segmentation in CT images with multimodal U-net  
智能系统学报. 2018, 13(6): 981–988 <https://dx.doi.org/10.11992/tis.201806012>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201908015

# 基于视-触跨模态感知的智能导盲系统

朱文霖<sup>1</sup>, 刘华平<sup>2</sup>, 王博文<sup>1</sup>, 孙富春<sup>2</sup>

(1. 河北工业大学 省部共建电工装备可靠性与智能化国家重点实验室, 天津 300130; 2. 清华大学 智能技术与系统国家重点实验室, 北京 100084)

**摘要:**盲人活动援助是盲人日常生活的重要组成部分。这些技术大多用于帮助盲人导航和躲避障碍物, 很少有研究将地面信息转换成一种给用户直观感受的触觉信息。为了满足上述需求, 本文提出了一种可以提供触觉反馈的盲人辅助地面识别智能导盲杖系统。试图利用深度生成对抗训练的方法来产生振动触觉刺激, 使用改进的 DiscoGAN 训练了我们的端到端生成网络。为了训练我们的网络, 构建了视触跨模态数据集 GroVib。通过上机实验和实物实验来评估方案的可行性, 通过上机实验结果表明参与者通过触觉识别地面的准确率为 84.7%, 触觉的平均真实感受得分为 71.3, 在真实场景实验中, 参与者只需平均 3.35 次尝试就可以根据触觉反馈来识别地面。

**关键词:**盲人用户; 电子手杖; 跨模态技术; 触觉; 数据集; 深度学习; 计算机视觉; 生成对抗网络

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2020)01-0033-08

中文引用格式: 朱文霖, 刘华平, 王博文, 等. 基于视-触跨模态感知的智能导盲系统 [J]. 智能系统学报, 2020, 15(1): 33-40.

英文引用格式: ZHU Wenlin, LIU Huaping, WANG Bowen, et al. An intelligent blind guidance system based on visual-touch cross-modal perception[J]. CAAI transactions on intelligent systems, 2020, 15(1): 33-40.

## An intelligent blind guidance system based on visual-touch cross-modal perception

ZHU Wenlin<sup>1</sup>, LIU Huaping<sup>2</sup>, WANG Bowen<sup>1</sup>, SUN Fuchun<sup>2</sup>

(1. State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin 300130, China; 2. State Key Lab. of Intelligent Technology and Systems, Tsinghua University, Beijing 100084, China)

**Abstract:** Blind mobility aid is a primary part of the daily life for blind people. Most of these technologies are used to help them navigate and avoid obstacles, and few researches have been done to convert ground information into tactile sensation that gives the user an intuitive feeling. To meet the above requirements, we proposed an intelligent guided rod system, which can provide tactile feedback to assist the blind to recognize ground information. We attempted to generate the vibrotactile stimuli leveraging the power of deep generative adversarial training. Specifically, we used an improved DiscoGAN training an end-to-end generated network. To train the network, we also built GroVib, a visual touch cross-modal dataset. We set up computer experiments and physical experiments to evaluate the feasibility of the scheme. The results of computer experiments showed that the accuracy rate of the participants in recognizing the ground by tactile sense was 84.7%, and the average real feeling score of tactile sense was 71.3. In real scene experiments, the participants needed only 3.25 times of attempts on average to recognize the ground based on tactile feedback.

**Keywords:** blind users; electronic cane; cross-modal technology; touch; data set; deep learning; computer vision; GANs

收稿日期: 2019-08-21.

基金项目: 国家自然科学基金重点项目 (U1613212); 河北省自然科学基金项目 (E2017202035).

通信作者: 刘华平. E-mail: hpliu@tsinghua.edu.cn.

视力障碍是困扰人们的最严重的问题之一。根据 WHO<sup>[1]</sup> 的报告称 80% 的眼睛疾病是可以预防甚至治愈的, 但失明仍然是一个严重的世界性

问题。除了努力在医学、神经科学和生物技术方面为这些问题找到最终的解决方案外,科学技术还可以通过提供基本功能(如独立导航和感知周围环境的能力)来为这些人提供帮助。

虽然目前的盲人辅助技术提供了一些先进的方案来解决盲人的行动指导或识别问题,但大部分工作对于使用者的反馈集中在语音告知或提供简单的定向引导振动。然而,触觉对视障患者来说是一种非常敏感的感知方式。将触觉信号引入盲人辅助中,可以使盲人用户更加直观地辨别地面环境,丰富盲人感知现实世界的方式,对于构建残疾人友好社会具有重要的应用价值。目前仍然缺乏从环境信息中产生触觉振动的解决方案。

触觉振动反馈可以使人类通过工具与物体表面的相互作用来感知纹理的属性。为了给人类提供这样的响应,触觉振动模型的研究已经开展了很长时间。遗憾的是,目前还没有基于工具状态和表面状态交互产生振动触觉响应的模型。这样的模型需要学习大量复杂的输入域和输出域空间映射:输入为工具状态(如工具速度),纹理表面状态(如地面纹理属性),输出为振动触觉信号。并且传统的振动信号生成模型只支持单一类型信号的输出。而这样的模型是无法满足现实生活中多种多样地面类别的需求的。

最近,从任一高维分布中生成新样本的生成方法得到了广泛的应用。具体地说,生成对抗网络(GAN)<sup>[2]</sup>在合成真实世界图像方面显示出了出色的能力。之前的研究表明,GANs可以有效地生成标签<sup>[3]</sup>、文本<sup>[4]</sup>等条件下的图像。尽管这些结果非常的出色,但很少有研究使用GANs来建模时间序列数据分布。

现代深度学习技术对于图像特征的提取和高维分布的拟合是非常依赖数据集的。现有的大多数公共数据集集中在体现视觉单一模态图像的差异性中。触觉和视觉属于两种模态的信息,目前尚未有直接可用的视触跨模态图像数据集。

本文的主要贡献有3个方面:

1) 介绍了盲人辅助设备的发展在此基础上,设计了一种面向视障人群的触觉反馈智能导盲杖装置。

2) 设计了一种基于GAN的端(地面信息)到端(振动反馈)的生成网络结构。

3) 构建了基于地面图像和振动信息的跨模态数据集(GroVib)。

## 1 相关工作

盲人辅助技术在过去的十几年中得到了突飞猛进的发展。研究结果主要针对导航辅助和识别辅助两个方面。文献[5-6]设计了一款可穿戴智能服装原型,传感器探测所关注的区域,并提供导航提示。文献[7]为盲人提供了一种基于超声波的导航辅助设备,使盲人能够通过安装在肩膀上的超声波传感器和一根导盲杖在前方6 m内探索路线。针对盲人的纸币识别已经在文献[8]中得到了解决,其中使用了加速鲁棒特性(SURF)。文献[9]处理了一个超市购物场景。在这项工作中,射频识别(RFID)被用作定位和导航的手段,而产品识别则是通过便携式相机读取二维码来实现的。产品条形码的检测和读取也在文献[10]中被解决。文献[11]提出了一种基于便携式摄像机的公交线路号码检测系统的设计方案。文献[12]提出了室内楼梯检测方案。文献[13]提出了一个通过内置的盲文脚本为盲人阅读电子书提供便利的设备。文献[14]认为服装颜色以及模式识别是促进盲人识别能力的一种手段。文献[15]提出了利用光学字符识别功能来识别文本,提供语音反馈的辅助文本阅读系统。

## 2 系统设计

触觉再现装置通常是通过各种方法来刺激皮肤的设备,如空气波纹管、微针阵列、直流电脉冲器和功能性神经肌肉刺激器等。本文的触觉输出模块是一种音频振动器,属于电励磁产生的振动装置。其优点是采用电机作为执行机构,可以获得任意方向的扭矩,响应速度相对较快,对不同频率的输入信号非常敏感。该装置可以模拟原始触觉信号采集的工具交互状态(如加速度等),这是选择该装置作为触觉输出的主要原因。

处理器本文选取了华为公司生产的Hikey970智能芯片,它是华为第三代开发板,具有更强的计算能力、更丰富的硬件接口,支持主流操作系统和人工智能模块。HiKey970集成了华为创新设计的HiAI框架,以及其他主流的神经网络框架,不但支持CPU、GPU的AI运算,还支持基于NPU的神经网络计算硬件加速,能效和性能分别可达CPU运算的50倍、25倍。本文将它烧录为Debian系统并在上面运行Tensorflow框架下的深度学习网络。摄像机选用一款拥有200万像素的USB摄像头模块,该款相机支持逆光拍摄且拍摄广角无畸变。处理器的音频输出功率相对较低,



不足以驱动振动模块, 因此选取了 LEPY 公司生产的 LP-2020A 型号的功率放大器将声波信号进行放大。功率放大器以及为整个系统提供电能的移动电源均放置在用户的背包中。振动块, 处理器, 相机以及控制按钮都集成在设计的 3D 打印的导盲杖中。模型的整体结构展示如图 1 所示, 图 2 展示了该设备的使用流程。



图 1 系统原型

Fig. 1 The prototype of the proposed system

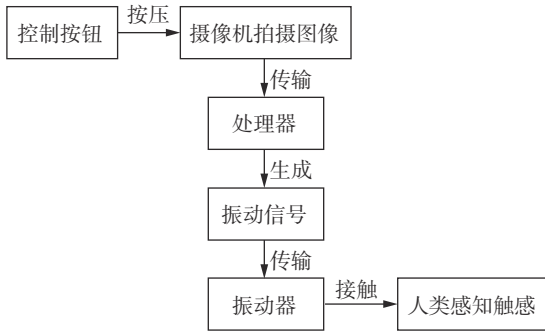


图 2 使用流程

Fig. 2 Gray scale image of sensor

### 3 系统实现

该系统的主要目标是自动生成可以反应目标地面特征的频谱图。在计算机视觉和机器学习领域, 这个问题可以看作图像到图像的翻译问题。文献 [18] 提出了一种识别输入图像类别, 然后输出相应类别信息的方法。然而, 这种方法将图像分类为类别标签, 会丢失很多高纬度特征信息, 因此本文需要更直接的解决方案。

GAN 是一种用于样式转换以及图像自动生成的方法。特别的, 基于 GAN 的方法可以用于在相同域或是跨域样式转换的生成。为了在不受用户干扰的情况下将地面图像转换成频谱, 采用

了跨域生成对抗网络作为该方法的基本组成部分。并在此基础上建立了一个端到端完整的生成网络。

#### 3.1 原始 GAN

GAN 是一种在无监督情况下生成逼真图像的有效方法。GAN 利用生成网络  $G$  将一个潜在的噪声变量  $z \sim P_{\text{noise}}$  转换为一个伪图像样本  $G(z)$ 。原始的 GAN 使用最小-最大博弈策略来训练生成器  $G$ , 并使用另一个网络  $D$  来区分  $G(z)$  和真实样本。形式上, GAN 的目标可以表示为

$$\min_G \max_D L(D, G) = E_{x \sim P_{\text{data}}(x)} [\log D(x)] + E_{z \sim P_{\text{noise}}} [\log (1 - D(G(z)))]$$

式中, 鉴别器  $D$  试图最大化真实样本输出的置信度得分。同时, 它也最小化了  $G$  生成的伪样本输出的置信度得分。相比之下,  $G$  的目标是最大化  $D$  对其输出的评价得分, 这可以看作是对  $D$  的一种欺骗。

#### 3.2 基于 DiscoGAN 的图像翻译

我们的任务是将地面图像域  $A$  转换为频谱图像域  $B$ 。解决这个问题最流行的方法是 Pix2Pix<sup>[19]</sup>、CycleGAN<sup>[20]</sup> 和 DiscoGAN<sup>[21]</sup>。其中 Pix2Pix 结合 GAN 损失和 L1 正则化损失, 提出了一种新的图像生成方法。然而, 这种方法需要一组配对的图像来监督。文献 [20-21] 提出了无监督图像到图像的生成方法以克服这一局限性。在这些方法中, 每个域都有一对生成器和鉴别器。对于给定的域  $A$  到  $B$ ,  $G_{AB}$  和  $G_{BA}$  将图像从  $A$  传输到  $B$  (或  $B$  到  $A$ ), 同时鉴别器  $D_A$  和  $D_B$  判断样本是否属于该域。CycleGAN 和 DiscoGAN 都提出了一个循环框架, 将图像从  $A$  域转换为  $B$  域, 然后再将其转换回原域, 从而在两个域之间建立强映射。为此, 这些方法的损失函数在这两个域之间有一个对抗性损失和一个循环 (重建) 损失, 定义如下:

$$\begin{aligned} L(A \rightarrow B) &= E_b [\log (D_b(b))] + \\ &E_a [\log (1 - D_b(G_{AB}(a)))] + \\ &E_a [\|G_{BA}(G_{AB}(a)) - a\|_2] \\ L(B \rightarrow A) &= E_a [\log (D_a(a))] + \\ &E_b [\log (1 - D_a(G_{BA}(b)))] + \\ &E_b [\|G_{AB}(G_{BA}(b)) - b\|_2] \end{aligned}$$

在本工作中, 由于 CycleGAN 有一个不适合我们问题的特点, 即它试图保留原始图像域的轮廓, 所以本文的方案最终采用了 DiscoGAN。图 3 展示了系统的整体框架。

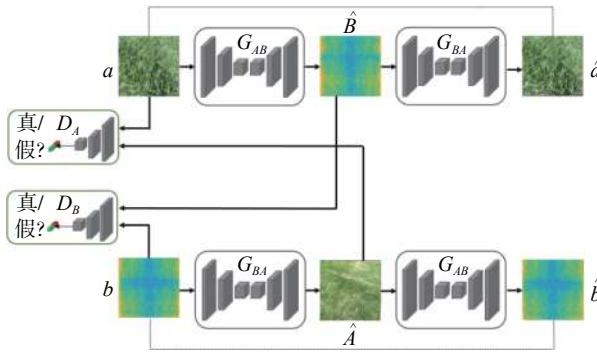


图 3 结构框架

Fig. 3 The framework of our proposed structure.

### 3.3 端到端生成网络

图 4 对数据流进行了逐级描述。当输入地面图像时,它将通过训练好的生成网络  $G$ 。 $G$  将它

转换成频谱图,最后,利用 Griffin-Lim 算法<sup>[22]</sup>将生成的谱图转换为声波格式。然后通过放大器和振动器将波形数据传输给用户。图 5 给出了 6 种场景振动信号波形的差异。

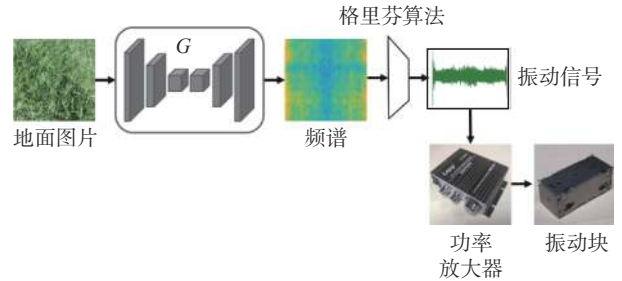


图 4 端到端生成模型

Fig. 4 The end-to-end Generate model

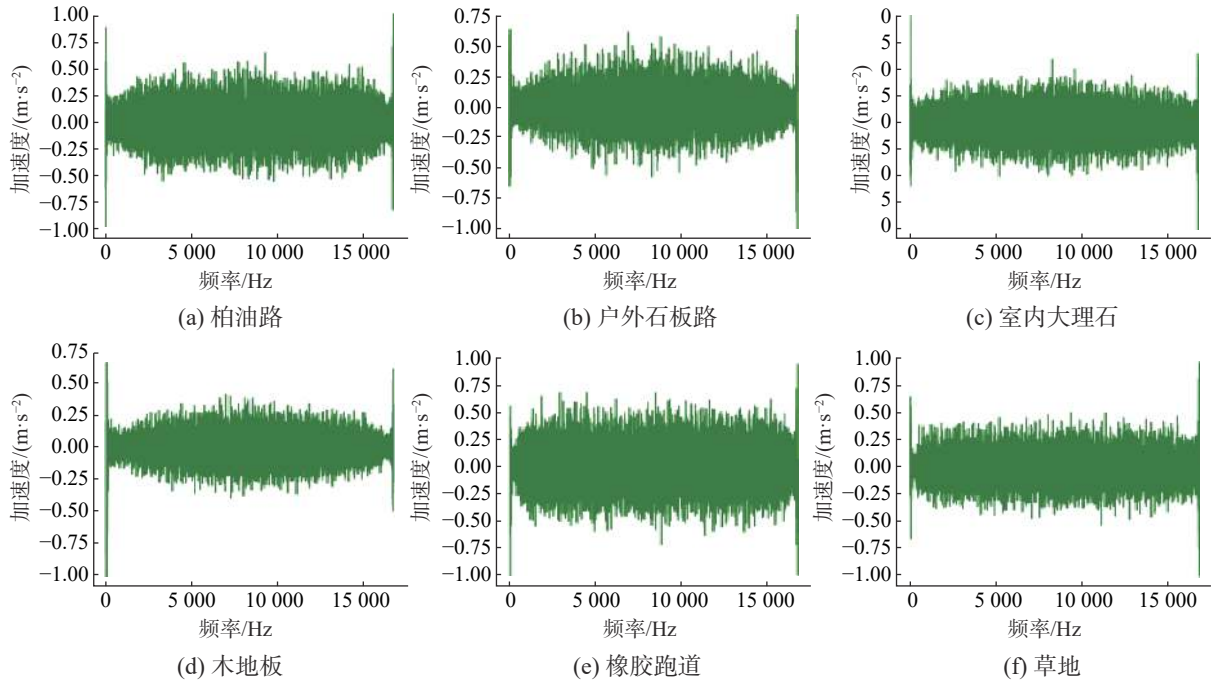


图 5 振动信号的波形展示

Fig. 5 Tactile Signal Waveform Display

## 4 GroVib 数据集

视觉和触觉是人类两种重要的感知方式,它们之间有着很强的联系,盲人可以通过触觉感知来弥补视觉的缺失。因此,构建了一个视触交叉模态数据集 GroVib(ground-vibration)作为深度学习网络训练的基础。

视觉图像易于捕捉,所包含的信息也很明显。但振动信息无法直观地描述。当工具敲击或在物体表面拖动时,会产生振动,振动可以通过加速度传感器捕捉到,相应的表面信号代表了物体表面的特征<sup>[23]</sup>。为了便于处理这些数据,将加

速度时间序列信号转换为图像。

GroVib 数据集由 6 种常见地面类型组成:柏油路、户外石板路、室内大理石、木地板、橡胶跑道和草地。

### 4.1 地面图像

数据集中的地面照片是通过智能手机(小米 8)进行拍摄的,拍摄的高度与导盲杖中相机所处位置相当(距地面约 80 cm)。为了保证模型可以全天候的使用,对高光照强度和昏暗条件下都进行了数据的采集,两种光强下对每一种地面各拍摄了 100 张照片,因此 GroVib 数据集集中有 1 200 张地面图像。

现代智能手机的拍照质量越来越高, 而我们的任务并不需要使用如此高分辨率的图片来完成, 所以需要将收集到的图片统一到一个较小的尺寸。我们裁剪了 1 200 张图片, 并将它们调整为  $256 \times 256$  的大小。图像的缩小不建议简单地缩小尺寸, 这样会造成严重的失真。因此, 使用 TensorFlow 中提供的面积插值方法对图像进行缩放。地面图像实例如图 6 所示。

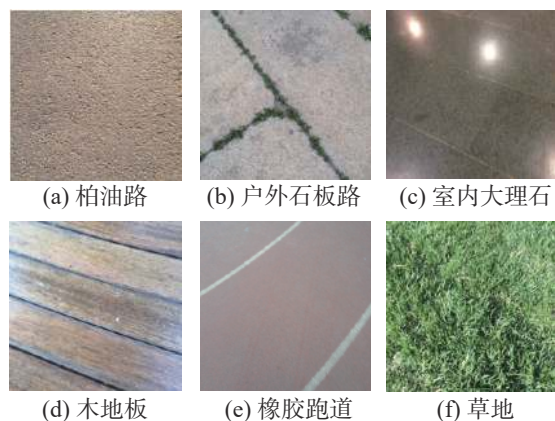


图 6 地面图像实例

Fig. 6 Gray scale image of sensor

## 4.2 振动图像

在本文的模型中, 加速度信号被用作振动触觉刺激。加速度信号参考了公开数据集 LMT-108-Surface Materials-Database<sup>[24]</sup> 中的结果, 该数据集包含了各种类型的纹理图像和相应工具产生的用于滑动任务的加速度信号。每种类型有 20 组加速度信号样本, 每组包含  $X$ 、 $Y$ 、 $Z$  等 3 个时间序列的信号。这些信号和图像均用 108 个类标签进行了标注。

从 108 类中, 选择 6 个小类对应于设置的地面类型: G2Brick 对应柏油路, G2StoneTileVersion2 对应户外石板路, G2Marble 对应室内大理石, G4CherryTree 对应木地板, G5ProfiledRubberPlate 对应橡胶跑道, G6FineArtificialGrassFibers 对应草地。

目前, 很少有研究使用 GANs 生成时间序列数据。这是因为 GANs 虽然擅长生成二维图像, 但在生成时间序列数据方面却很差。因此, 选择频谱图作为加速度信号的表示。原始数据库中每个信号有 4 s 长, 采样率为 10 kHz。本文使用短时傅里叶变换 (STFT) 完成信号的转换步骤。STFT 用于确定时变信号局部区域正弦波的频率和相位, 其处理的信号具有时域和频域的局域特性。在这一步中, 对变换的结果进行对数运算和归一化运算, 并且提取 0~1.625 s, 频率在 0~256 Hz

的频域内的信息。最后裁剪了频谱图, 把它们调整成到  $256 \times 256$  的大小。短时傅里叶变换的具体参数如表 1 所示。生成的频谱样式如图 7 所示。

表 1 STFT 的参数  
Table 1 STFT parameters

| 参数名称    | 参数数值   |
|---------|--------|
| Fs      | 10 000 |
| 重叠      | 128    |
| 时间/s    | 1.625  |
| 汉明窗     | 512    |
| nfft    | 512    |
| 频率范围/Hz | 0~256  |

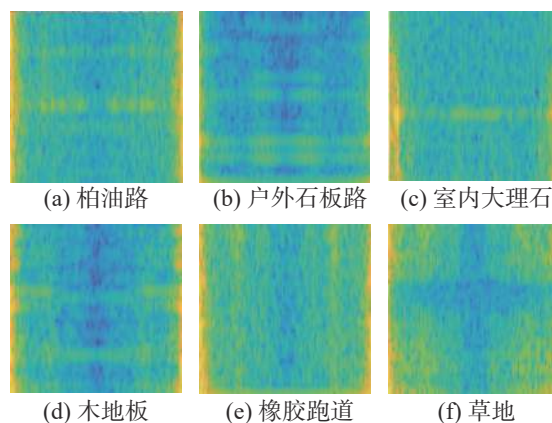


图 7 频谱图像实例

Fig. 7 Gray scale image of sensor

## 5 生成结果

### 5.1 实验装置

我们的模型是使用一台配备英特尔酷睿 i7-8750H CPU、16 GB DDR4 RAM 和 NVidia Geforce GTX 1080 Ti 的台式电脑完成的。

### 5.2 数据预处理

虽然本文 GroVib 数据集对于每一类地面都有 200 张照片, 但是对于训练一个神经网络来说仍然是不够的, 所以在训练模型时遵循数据增强设置。将每张图像进行水平和垂直翻转, 以及随机旋转它们的角度, 除了传统方法外, 还使用随机擦除和混合技术。将每个类别的地面数据扩展到 1 000 个, 因此总共有 6 000 个大小为  $256 \times 256$  的地面图像。而频谱图不能用于数据增强, 因为它们代表了振动信号的时间序列特征和振幅强度, 类似的处理方法会严重影响隐藏信息。本文对地面图像和频谱图都进行了标注, 这样在训练模型时可以根据类别信息选择相应的数据。



### 5.3 生成结果

从实验结果图8可以看出,本文方法可以根据地面图像的输入生成相应类别的频谱图,生成效果可以既保证良好的多样性,同时又不失很强的相似性。从不同轮次生成的结果可以看出,10个轮次之后生成的图像与真实的频谱已经具有一定的相似性,但细节模糊,粒度感较强。经过200个轮次的训练,生成结果有较高的清晰度,经过400个轮次的训练,生成的频谱与真实数据有很高的相似性,所以在400个轮次之后停止了网络训练。

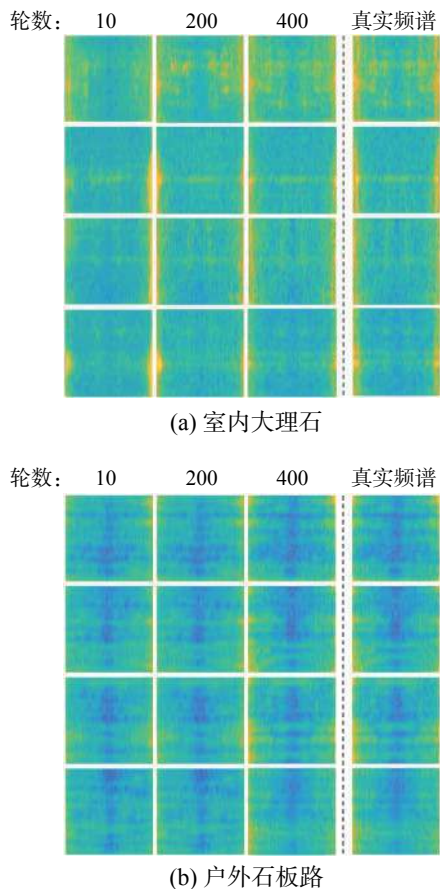


图8 不同轮次频谱生成的结果

Fig. 8 The result of generated spectrograms with different epochs

## 6 用户测试

在实际场景中评估了所提出系统的性能。首先,测试了方法是否能够生成近乎真实的振动触觉刺激,以及用户是否能够区分出振动所属的类别。然后,进行了另一项用户研究,以评估在真实场景中模拟盲人用户正确识别地面信息所需尝试的次数。10名年龄在21~26岁之间的志愿者(7名男性和3名女性)参与了这些研究。研究人员对他们进行了筛选,以确定他们没有抑郁,也没有极度疲劳,因为他们的知觉会受到身体或情

绪状态的影响。

### 6.1 振动结果感受实验

产生具有强烈真实感的振动信号对提高用户体验具有重要意义。在测试过程中,受试者被蒙住眼睛,用惯用手握住导盲杖的把手部分(图9)。研究人员在测试界面上选择信号的类别以及是真实还是生成的数据类型。在产生振动之后,参与者将回答3个问题:他们感觉到的是什么类别的触感,这种振动是真实的还是生成的。此外,他们还通过视觉模拟量表(VAS)<sup>[25]</sup>来描述每个刺激的真实度。参与者在这个测试方法的模拟尺度上对他们的真实感受进行评分。他们将要描述的是“该触觉感觉有多么的真实?”参与者通过观察和记忆实验前的一条9等分线段(图9),其中0表示他们感觉到的振动是随机的、不真实的,10表示振动非常接近真实的感觉。每个参与者在每个类别中测试5个真实的振动和5个生成的振动。结果总共记录了600个测试结果。为了防止顺序效应,这些类别的呈现顺序被随机分配,并在参与者之间进行平衡。



图9 实验硬件框架

Fig. 9 Experimental hardware diagram

如图10所示,参与者平均识别精度为84.7%,这些结果表明振动信号具有足够的可分辨性,用户可以识别出地面类别,证明了通过触觉感知地面信息的可行性。图11(a)显示了正确识别是哪个刺激产生的百分比。我们把这个值称为正确答案率。如果这个值接近50%,则意味着参与者无法区分真实数据和生成的数据。在问卷调查后,所有参与者都回答说,他们没有发现真实的刺激和生成的刺激之间的区别。因此,可以说我们的系统有潜力产生高逼真的振动触觉信号。另一方面,图11(b)显示了参与者对每个类别的真实感受。我们可以直观的看出,生成的数据与真实数据的没有显著的差异,它们平均真实感得分为71.3,生成数据的平均真实感得分为大多在70分以上。

结合这些实验结果可以看出,草地的识别精

度最高,但真实性得分却很低。反之,识别结果处于中等水平的室外石板路获得的真实感得分最高。因此,真实度评价与生成数据的识别率之间不存在相关性。

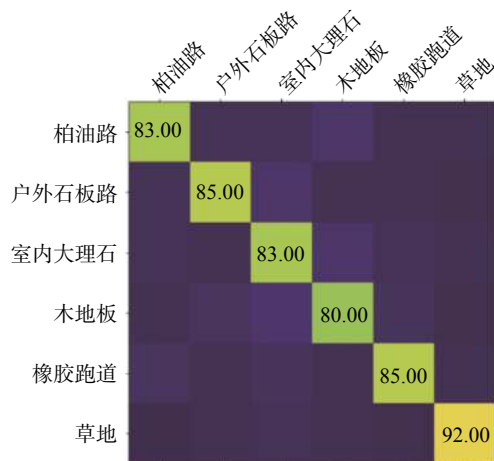
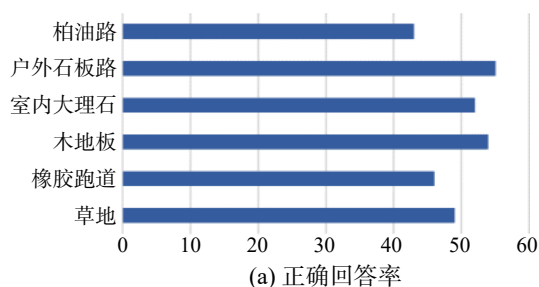
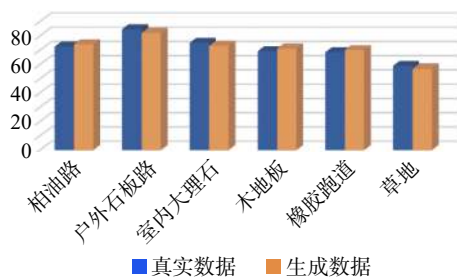


图 10 触觉识别结果的混淆矩阵

Fig. 10 Confusion matrix of tactile recognition results



(a) 正确回答率



(b) 参与者感受到的真实感评分

图 11 参与者触觉感受的结果

Fig. 11 The results of the participants' tactile perception

## 6.2 真实使用场景模拟实验

我们蒙住所有志愿者的眼睛,让他们戴上隔音耳塞,因此他们听不到环境中动态障碍物产生的噪音,比如在柏油路上行驶的汽车,这样他们就不能通过联想来判断地面。在每个场景中,参与者都有足够的活动空间,为了保证志愿者的安全,每组实验的参与者都有研究人员的指导。还有一件需要考虑的事情是,由于视觉正常的参与者被带到某个场景中会产生记忆效应,严重影响测试的准确性,所以他们必须被研究人员牵引到一个陌生的环境中再进行实验。参与者手持导盲

杖,按下控制器拍下地面的照片,然后根据他们感觉到的触觉振动回答他们的识别结果。当参与者做出正确的判断时,实验结束,研究人员记录下他们回答的次数。真实场景仿真实验结果如图 12 所示,正确识别平均所需次数为 3.35 次,其中正确识别草地所需的尝试次数最少。这是因为人们草的触感的分辨率最高。这一结论与前面的实验结果是一致的。

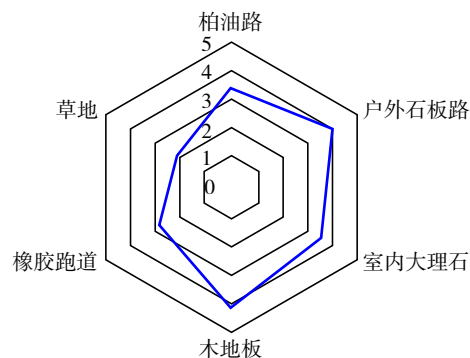


图 12 参与者做出正确识别所需的平均次数

Fig. 12 The average number of times participants needed to make a correct recognition

## 7 结束语

本文介绍了一种为视障人士提供触觉识别辅助的导盲杖设备。为了训练模型,我们建立了一个跨模态数据集 GroVib,将时间序列信号转换成二维图像进行网络处理。本文利用 GAN 的思想对振动信号的生成网络进行训练。实验结果表明,用户使用我们的导盲杖系统可以仅经过 3 次左右的尝试就可以识别出所处地面的信息,该系统可以为盲人用户带来更好的旅行体验。本文的研究朝着更全面的盲人辅助技术迈出了重要的一步。

## 参考文献:

- [1] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005:539-546
- [2] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA, 2014: 2672-2680.
- [3] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. arXiv: 1411.1784, 2014.
- [4] REED S, AKATA Z, YAN Xinchun, et al. Generative adversarial text to image synthesis[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York, USA, 2016: 1060-1069.



- [5] BAHADIR S K, KONCAR V, KALAOGLU F. Wearable obstacle detection system fully integrated to textile structures for visually impaired people[J]. *Sensors and actuators A: physical*, 2012, 179: 297–311.
- [6] SHIN B S, LIM C S. Obstacle detection and avoidance system for visually impaired people[C]//Proceedings of the 2nd International Workshop on Haptic and Audio Interaction Design. Seoul, South Korea, 2007: 78–85.
- [7] BOUSBIA-SALAH M, BETTAYEB M, LARBI A. A navigation aid for blind people[J]. *Journal of intelligent & robotic systems*, 2011, 64(3/4): 387–400.
- [8] HASANUZZAMAN F M, YANG Xiaodong, TIAN Yingli. Robust and effective component-based banknote recognition for the blind[J]. *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, 2012, 42(6): 1021–1030.
- [9] GUEST S, DESSIRIER J, MEHRABYAN A. The development and validation of sensory and emotional scales of touch perception[J]. *Attention perception & psychophysics*, 2011, 73(2): 531–550.
- [10] KIM D Y, YI K Y. A user-steered guide robot for the blind[C]//Proceedings of 2008 IEEE International Conference on Robotics and Biomimetics. Bangkok, Thailand, 2009: 114–119.
- [11] TIWANA M, REDMOND S, LOVELL N. A review of tactile sensing technologies with applications in biomedical engineering[J]. *Sensors and actuators: a physical*, 2012, 179(5):17–31.
- [12] TANG T J J, LUI W L D, LI W H. Plane-based detection of staircases using inverse depth[C]//Proceedings of 2012 Australasian Conference on Robotics and Automation. New Zealand, 2012: 1–10.
- [13] AL KALBANI J, SUWAILAM R B, AL YAF AI A, et al. Bus detection system for blind people using RFID[C]//Proceedings of the 2015 IEEE 8th GCC Conference & Exhibition. Muscat, Oman, 2015: 1–6.
- [14] KULKARNI A, BHURCHANDI K. Low cost E-book reading device for blind people[C]//Proceedings of 2015 International Conference on Computing Communication Control and Automation. Pune, India, 2015: 516–520.
- [15] THILAGAVATHI B. Recognizing clothes patterns and colours for blind people using neural network[C]//Proceedings of 2015 International Conference on Innovations in Information, Embedded and Communication Systems. Coimbatore, India, 2015: 1–5.
- [16] NICHOLLS H, LEE M. A survey of robot tactile sensing technology[J]. *The international journal of robotics research*, 1989, 8(3):3–30.
- [17] STRESE M, SCHUWERK C, IEPURE A, et al. Multimodal feature-based surface material classification[J]. *IEEE transactions on haptics*, 2017, 10(2): 226–239.
- [18] LI Xinwu, LIU Huaping, ZHOU Junfeng, et al. Learning cross-modal visual-tactile representation using ensemble generative adversarial networks[J]. *Cognitive computation and systems*, 2019, 1(2): 40–44.
- [19] ISOLA P, ZHU Junyan, ZHOU Tinghui, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 5967–5976.
- [20] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2242–2251.
- [21] KIM T, CHA M, KIM H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 1857–1865.
- [22] GRIFFIN D, LIM J. Signal estimation from modified short-time fourier transform[J]. *IEEE transactions on acoustics, speech, and signal processing*, 1984, 32(2): 236–243.
- [23] UJITOKO Y, BAN Y. Vibrotactile signal generation from texture images or attributes using generative adversarial network[C]//Proceedings of the 11th International Conference on Human Haptic Sensing and Touch Enabled Computer Applications. Pisa, Italy, 2018: 25–36.
- [24] HUANG G, WANG D, LAN Y. Extreme learning machines: a survey[J]. *International journal of machine learning & cybernetics*, 2011, 2(2): 107–122.
- [25] LEE K A, HICKS G, NINO-MURCIA G. Validity and reliability of a scale to assess fatigue[J]. *Psychiatry research*, 1991, 36(3): 291–298.

#### 作者简介:



朱文霖,男,1994年生,硕士研究生,主要研究方向为新型磁性材料与器件、触觉交互。



刘华平,男,1976年生,副教授,博士生导师,主要研究机器人感知、学习与控制,多模态信息融合。利用稀疏编码建立了机器人多模态融合感知与学习框架,在此基础上结合机器人的光学、红外、深度和触觉等不同模态信息开发了一系列多模态稀疏编码方法,并在移动机器人、灵巧机械臂等机器人平台上开展多模态感知融合的方法验证与应用。发表学术论文10余篇。



王博文,男,1956年生,教授,博士生导师,主要研究方向为磁致伸缩材料与器件、振动发电技术、磁特性测试技术。承担国家自然科学基金等项目8项(其中主持5项)、省部级科研项目10项(其中主持8项);河北省科学技术突出贡献奖和省科技进步三等奖各1项。获专利授权6项;出版专著2部,发表学术论文200多篇。