

面向遥操作电力作业的触觉生成技术

刘跟创¹, 李 昂^{1,2*}

¹南京邮电大学通信与信息工程学院 江苏南京

²南京邮电大学空天地海通信技术一体化研究院 江苏南京

【摘要】融合触觉反馈的远程遥操作电力作业已成为提升高压环境作业安全性与操作精度的关键技术,但受限于触觉通信的带宽不足与信号易失真的问题。为此,本研究提出一种基于大模型的文本引导触觉生成技术,通过将高维触觉图像压缩为低维文本语义特征进行传输,显著降低了数据量,有效解决了带宽受限问题;文本特征的语义抽象与结构化编码机制显著提升了抗干扰性能,有效解决了传统触觉图像信号在传输过程中易失真的问题。

【关键词】遥操作电力作业; 文本特征提取; 触觉图像生成

【基金项目】国家自然科学基金(62306151)

【收稿日期】2025 年 3 月 6 日

【出刊日期】2025 年 4 月 6 日

【DOI】10.12208/j.jeea.20250107

Tactile generation technology for remote operation of power operations

Genchuang Liu¹, Ang Li^{1,2*}

¹*School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu*

²*Institute of Space-Air-Ground-Sea Integrated Communication Technology, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu*

【Abstract】 Remote teleoperation of power operations with haptic feedback has become a key technology for enhancing safety and operational accuracy in high-voltage environments. However, it is limited by insufficient bandwidth for haptic communication and the susceptibility of signals to distortion. To address these issues, this study proposes a text-guided haptic generation technology based on large models. By compressing high-dimensional haptic images into low-dimensional text semantic features for transmission, it significantly reduces the data volume and effectively resolves the bandwidth limitation problem. The semantic abstraction and structured encoding mechanism of text features significantly enhance the anti-interference performance, effectively solving the problem of signal distortion during the transmission of traditional haptic images.

【Keywords】 Teleoperation of power operations; Text feature extraction; Tactile image generation

引言

随着通信技术发展,利用遥操作技术实现远程电力作业已经成为保障电力系统安全稳定运行、提高作业效率的重要手段。遥操作技术是指借助信息化手段,通过交互设备,机器人接收指令并执行相应任务,同时将自身状态和所处环境信息反馈给人类,从而实现人类对机器人远程实时控制的技术。遥操作技术通过传感器和自动化设备替代人工电力作业,避免人员直接接触高压设备、高温环境或复

杂地形带来的安全风险。

早期遥操作技术以视觉反馈为主导,通过远程控制机器人执行巡检任务,Elizondo 等最早提出将机器人应用于带点作业的概念框架^[1]。随着多模态技术领域不断发展,宋爱国团队系统提出多模态力触觉交互技术理论框架^[2],将视觉反馈与力触觉反馈联合作为操作员的判断依据,在这之后,Jing Zhu 等创新性地将 AR 技术与力反馈结合^[3],提出图形力反馈(GFF)系统。在人工智能技术发展迅速的现代,

*通讯作者: 李昂

Suresh 等开发 NeuralFeels 视触融合方法^[4], 利用神经网络将视觉与触觉的多模态信息进行融合远程传输, 完成基于双模态信息反馈的遥操作电力作业。

但现有集成触觉通信的遥操作技术存在一定的技术壁垒, 一方面, 有限的传输带宽使得力触觉反馈的传输无法满足高速率, 低时延的通信需求; 另一方面, 触觉信号在传输过程中易受电力作业环境强电磁干扰等的影响, 导致触觉数据失真, 进而降低操作者对远程设备触觉感知的准确性。

针对上述问题, 本研究提出以文本模态作为媒介引导触觉图像生成的技术路径。从理论机制来看, 文本模态具备天然的数据轻量级特性与抗干扰特性: 一方面, 在各类模态中, 文本模态的数据量最小, 这一特性使其在数据传输方面具备显著优势; 另一方面, 文本的语义抽象和结构化编码机制, 显著提升了数据传输过程的抗干扰性能。这一机制为解决遥操作电力作业中的触觉通信瓶颈提供了新思路。

在文本与触觉模态关联的任务中, Fengyu Yang 等人提出的 Touch-LLM 系统通过跨模态对齐技术, 实现了大型语言模型 (LLM) 对触觉的语义理解能力, 揭示了文本与触觉模态的深度关联^[5]。这一突破表明, 大模型通过语义空间映射, 能够将触觉的物理刺激与文本的抽象概念建立可计算的对应关系。

由此, 本文将探讨一种新的技术方案, 通过利用大模型语义理解能力和文本-触觉跨模态通信技术, 提出了一种基于大模型的文本引导触觉生成技术。具体的工作如下:

(1) 该系统通过文本特征传输与文本引导触觉图像生成相结合的方式替代触觉图像直接传输, 显著降低了通信带宽需求, 有效缓解了远程传输的数据压力; 另一方面, 基于文本语义特征的数据传输机制对电力作业环境中的干扰具有鲁棒性 (robustness), 确保了传输过程中数据完整性和准确性, 大幅提升了触觉反馈的可靠性。

(2) 通过有监督学习与迁移学习相结合的方法, 设计了分阶段三层训练策略, 逐步实现触觉图像的有效生成, 并有效降低了计算需求。

(3) 将触觉图像生成器制作成应用程序接口, 并设计了用户交互平台, 增强人机交互性的同时, 增加了系统的可调用性。

1 基于大模型的文本引导触觉生成

1.1 算法设计

本研究提出了文本引导触觉图像生成的算法实现方案, 具体如图 1 所示:

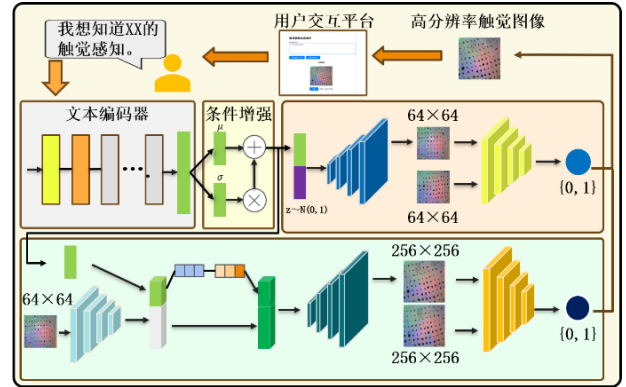


图 1 模型算法流程图

(1) 构建高维文本语义特征至触觉特征的关联实现跨模态对齐^[7]: 为弥合文本与触觉图像的模态鸿沟, 研究采用轻量级大模型 LLaMA-1B 的文本编码组件, 提取文本的高维语义特征, 与传统的循环神经网络构造的编码器对比, 大模型更全面地捕获文本中潜藏的材料属性。

$$\varphi_t = \text{LLAMA_Decoder}(t) \quad (1)$$

完成文本编码后, 通过条件增强模块^[7]: 对文本嵌入进行非线性变换, 经过感知层生成条件潜变量, 得到服从特定分布的隐变量。

$$\hat{c} = \text{CA_net}(\varphi_t) \quad (2)$$

具体过程则是从独立高斯分布 $N(\mu(\varphi_t), \Sigma\varphi_t)$ 中随机采样潜在变量 $\hat{c}$ 。其中平均值 $\mu(\varphi_t)$ 和对角协方差矩阵 $\Sigma\varphi_t$ 均为文本嵌入 φ_t 的函数。

基于大模型的文本特征提取, 能够更全面地提取完整的文本语义, 而条件增强这一操作通过特征形式转化策略缩小文本特征 φ_t 与触觉特征的模态差异。使得该隐变量在潜空间中靠近真实触觉图像的分布区域, 从而实现初步引导生成器输出符合该材料特性触觉图像。

(2) 双层生成架构设计: 为了解决触觉图像分辨率与细节表征的瓶颈, 本研究提出以堆叠式生成对抗网络 (StackGAN) 为基础的双阶段生成框架, 依靠从粗到细的细节增强, 凭借分阶段优化达成触觉信号的多尺度生成, 第一阶段将基于轻量级大模型的文本编码器 LLaMA_Decoder 所提取高维语义特征 φ_t 转化来的高斯特征向量 $\hat{c}$ 当作条件输入, 引导第一层生成器生成低分辨率触觉图像:

$$TI_{(64 \times 64)} = G_0(z, \hat{c}_0) \quad (3)$$

z 是从给定分布 p_z (本文为高斯分布) 中随机采样的噪声向量, $\hat{c}_0$ 为第一层生成器 G_0 的输入条件向量, $TI_{(64 \times 64)}$ 为阶段一生成的低分辨率触觉图像。

此时的触觉图像具备材料的综合特征, 然而微观细节仍然模糊。紧接着, 将第一阶段的输出以及原始文本特征输入第二层生成器, 通过级联的上采样模块和残差注意力机制构成的生成器, 逐渐还原触觉图像的局部细节。

$$f_l = Ds(TI_{(64 \times 64)}) \quad (4)$$

f_l 是由阶段一生成的粗略触觉图像经由下采样模块 Ds 特征提取得来的低分辨率触觉图像特征。将 f_l 与噪声进行拼接, 得到阶段二生成器 G 的一路输入 s_0 , 另一路输入则为阶段二的高斯特征向量 $\hat{c}$ 。 $TI_{(256 \times 256)}$ 即为最终生成的触觉图像:

$$TI_{(256 \times 256)} = G(s_0, \hat{c}) \quad (5)$$

分阶段生成缓解了单阶段模型对高维触觉数据的直接建模困难。通过分层策略, 第二层网络可专注于细节增强, 避免全局-局部特征的优化冲突。

1.2 训练策略设计

为了进一步确保在解决触觉图像生成分辨率受限的同时保证触觉图像的生成质量, 本研究参照 StackGAN 的分阶段生成思想, 针对基于大模型的文本引导触觉图像生成任务, 提出有监督学习和迁移学习相结合的三级渐进式训练策略, 旨在实现从粗粒度全局特征到细粒度触觉图像的高保真生成, 同时降低计算需求。具体训练流程设计如图 2 所示:

第一阶段进行单材料低分辨率触觉生成训练, 以混凝土材料的触觉图像数据为基础, 初步培育文本语义到触觉图像的映射能力, 生成 64×64 的低分辨率触觉图像。先训练判别器, 再利用判别器的结果去引导生成器训练。判别器的损失函数如下:

$$\mathcal{L}_{D_0} = \mathbb{E}_{(I_0, t) \sim p_{data}} [\log(D_0(I_0, \varphi_t))] + \mathbb{E}_{z \sim p_z, t \sim p_{data}} [\log(D_0(G_0(z, \hat{c}_0), \varphi_t))] \quad (6)$$

生成器的损失函数如下: (1.1)

$$\mathcal{L}_{G_0} = \mathbb{E}_{z \sim p_z, t \sim p_{data}} [\log(D_0(G_0(z, \hat{c}_0), \varphi_t))] \quad (7)$$

其中, 真实图像 I_0 和文本描述 t 来自真实数据分布 p_{data} , z 是从给定分布 p_z (本文为高斯分布) 中随机采样的噪声向量, φ_t 为文本嵌入向量。

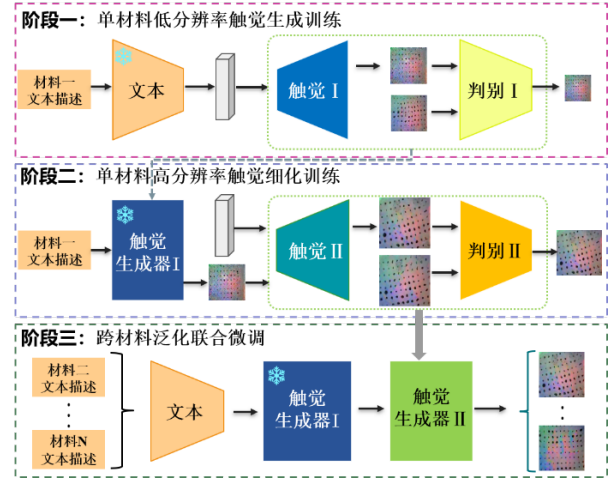


图 2 分阶段三层训练策略设计流程图

第二阶段进行单材料高分辨率触觉细化训练, 冻结阶段一生成器参数, 采取双通道输入训练阶段二生成器, 基于阶段一的低分辨率输出, 结合同一文本特征引导, 生成 256×256 高分辨率触觉图像, 补充微观触觉图像细节。训练第二层判别器的损失函数如下:

$$\mathcal{L}_D = \mathbb{E}_{(I, t) \sim p_{data}} [\log(D(I, \varphi_t))] + \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(D(G(s_0, \hat{c}), \varphi_t))] \quad (8)$$

生成器的损失函数如下:

$$\mathcal{L}_G = \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(D(G(s_0, \hat{c}), \varphi_t))] \quad (9)$$

与原始生成对抗网络公式不同, 在此阶段不使用随机噪声 z , 且假设随机性已由第一阶段低分辨率输出 s_0 保留。这一阶段使用的高斯条件变量 $\hat{c}$ 和阶段一网络使用的 $\hat{c}_0$ 来自于相同的文本嵌入 φ_t , 但阶段一和阶段二的条件增强模块具有不同的全连接层。通过这种方式, 阶段二生成器学习捕获阶段一生成器忽略的文本嵌入中的有用信息。

第三阶段进行跨材料泛化联合微调训练, 通过迁移学习和大模型微调的相关技术, 将阶段二生成器与 LLAMA 文本编码器在小规模多材料数据集上进行联合微调, 扩展模型对新材料类别的触觉生成能力, 实现跨材料泛化。采用低秩自适应 (LoRA) 轻量化的微调办法^[8], 完成文本—触觉跨模态表征的定向校准, 避免全参数微调出现的计算冗余。在联合微调阶段, 需要对不同组件的参数进行选择性冻结, 如表 1 所示:

表 1 联合微调阶段组件参数冻结情况表

组件	参数冻结	说明
StackGAN 第一层	√	保留预训练特征提取能力
StackGAN 第二层	×	仅微调高频细节生成模块
LLAMA 编码器	√	固定基础语言模型参数

LoRA 适配器主要注入 LLAMA 文本编码器的交叉注意力模块中, 选择注意力机制中的查询投影和值投影作为参数更新目标, 通过低秩矩阵分解实现参数高效调整。LoRA 的核心参数设置如下表 2 所示:

表 2 LoRa 核心参数设置

低秩矩阵维度	缩放系数	丢弃率
8	32	0.1

该设计使得模型仅需更新约 0.5% 的参数量, 就可实现触觉生成的跨材料泛化, 沿着梯度传播的路径, 文本特征先借助 LoRA 适配器进行语义重构, 随后作为条件输入注入到 StackGAN 第二层生成器里面, 从而构建了从文本语义空间到触觉特征空间的定向映射通道。

分阶段三层训练框架为模型提供了动态增量学习能力与知识继承机制, 有效支撑了模型的可扩展性。该技术路径支持触觉生成的持续性迭代开发, 通过增量式添加新类别数据结合迁移训练, 该模型可渐进式扩展至未识别的材料类型, 且不必改动模型架构, LLAMA 的跨领域语言理解支持其对开放指令的解析, 通过语义映射及触觉特征融合, 实现超出训练分布的新型触觉图像构建。

2 实验与验证

训练选用 Touch and Go 真实世界触觉图像数据集, 旨在生成与现实最接近的触觉感知。在触觉生成器预训练阶段, 选取混凝土材料的触觉图像作为有监督训练的单一材料数据集。

第一阶段进行第一层低分辨率触觉生成器训练。选用 Adam 优化器, 设置生成器学习率为 2×10^{-4} , 判别器学习率为 2×10^{-5} , 由于第一层生成对抗神经网络无法形成触觉图像细节特征, 设置 batch size 为 150 加速训练 (预训练数据集含 13, 000 张触觉图像), 共训练 20 个 epoch。

第二阶段进行第二层高分辨率触觉生成器训练。

选用 Adam 优化器, 设置生成器学习率为 2×10^{-4} , 判别器学习率为 2×10^{-5} , batchsize 为 32, 共训练 40 个 epoch。

至此, 基于 StackGAN 的触觉图像生成器的预训练已完成, 模型可以执行单材料的触觉图像生成。

第三阶段进行跨材料泛化联合微调训练, 冻结低分辨率触觉图像生成器, 训练文本编码器与高分辨率触觉图像生成器。选用 Adam 优化器, 设置文本编码器与高分辨率触觉图像生成器学习率为 1×10^{-4} , 为了防止出现模式崩溃, 设置判别器学习率为 5×10^{-6} , batchsize 为 32, 使用 accelerate 加速训练, 基于验证集损失曲线, 50 epoch 后模型性能趋于稳定。

针对高分辨率触觉图像生成器功能的验证, 本研究将混凝土与拓展材料作为两组测试样本: 其中混凝土是作为模型预训练的预训练用的单一材料类别。对于每种材料, 我们分别将多个描述其触觉特征的自然语言文本作为输入, 测试模型生成效果。

表 3 触觉图像生成 FID (弗雷歇特距离) 指标

测试组	混凝土 (单一材料)	拓展材料
生成质量 (平均 FID)	45.16	61.70

如表 3 所示, 混凝土材料触觉图像的生成结果与真实图像对比, FID 值为 45.16 (< 50), 模型在混凝土材料 (单一材料) 展现出了优秀的生成效果; 拓展材料的生成结果与真实图像相比, FID 值为 61.70 (< 100), 模型在拓展材料上展现出较好的生成效果。

实验证明, 基于轻量级 LLAMA 模型的文本引导触觉图像生成模型能够在实现计算效率优化的同时, 把握不同材料的关键触觉特性, 依靠文本引导触觉图像生成的方式实现触觉信息传输。

3 交互系统设计

整体交互平台采用分层架构, 通过标准化的数据流实现端到端的触觉生成服务。如图 3 所示, 基于文本的触觉图像生成系统的交互流程遵循清晰的端到端数据传输链路, 具体交互逻辑如下:

用户在前端界面的文本输入框里录入触觉描述, 然后点击“生成”按钮启动交互流程, 前端将用户所输入的文本内容封装成 JSON 格式请求体, 借助 HTTP POST 方法传至 Flask 后端服务的/generate 接口。Flask 后端在接收到请求后, 立即启动请求处

理流水线。首先对 JSON 数据进行完整性校验, 验证通过后, 后端通过进程间通信机制调用独立的触觉图像生成模型服务。该生成模型基于预训练的 LLAMA-StackGAN, 将文本语义编码为高维特征向量, 并通过生成对抗网络合成触觉矩阵数据。

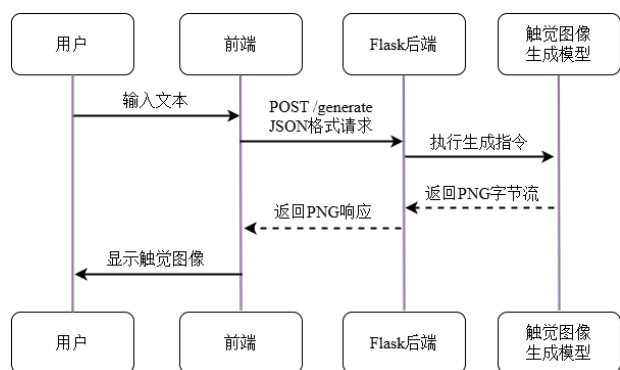


图3 前后端平台交互流程

触觉图像生成模型在完成计算后, 将输出的浮点矩阵数据转换成标准化的 PNG 图像字节流: 凭借归一化算法把触觉图像像素值映射至 0-255 的数值区间, Flask 后端将 PNG 字节流封装成 HTTP 响应, 并将响应头 Content-Type 设置为 png。系统采用分层通信协议栈保障数据传输效率与安全性。

4 结论

基于大模型的文本引导触觉生成技术通过跨模态语义压缩与重构机制, 创新性地解决了远程遥操作电力作业中的通信瓶颈问题, 压缩后的文本数据量既降低了通信带宽需求, 其离散化的文本特性对电力作业环境中的干扰具有天然鲁棒性, 因此, 理论上文本特征传输过程不易发生失真。本研究为触觉信息传输提供了新的技术方案, 利用 LLAMA 轻量级模型提取与触觉图像相关文本的高维文本语义特征, 将传统触觉数据流压缩为轻量级文本特征, 在接收端通过文本语义特征引导触觉图像生成, 该技术也为多模态通信方法提供了新的技术路径。

参考文献

- [1] Elizondo D, Gentile T, Candia H, et al. Overview of robotic applications for energized transmission line work — Technologies, field projects and future developments[J]. IEEE, 2010.
- [2] 张耀华, 宋爱国, 缪天缘, 等. 面向电力设施巡检运维的多模态遥操作机器人研究进展[J]. 中国测试, 2025, 51(04): 18-30.
- [3] Zhu J, Chen Y, Xu M, et al. Graphical Force and Haptic Feedback Teleoperation System for Live Power Lines Maintaining Robot[J]. IEEE, 2019.
- [4] Suresh S, Qi H, Wu T, et al. NeuralFeels with neural fields: Visuotactile perception for in-hand manipulation[J]. Science Robotics, 2024, 9(96).
- [5] Yang F, Feng C, Chen Z, et al. Binding Touch to Everything: Learning Unified Multimodal Tactile Representations[J]. IEEE, 2024.
- [6] Hu H, Chan K C K, Su Y C, et al. Instruct-Imagen: Image Generation with Multi-modal Instruction[J]. IEEE, 2024.
- [7] Zhang H, Xu T, Li H, et al. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks[J]. IEEE, 2017.
- [8] Hu E J, Shen Y, Wallis P, et al. Lora: Low-rank adaptation of large language models[J]. ICLR, 2022, 1(2): 3.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS