# 基于智能生成技术的手语数字人发展现状与趋势

□文 / 唐申庚 *，修雪玉，郭丹，洪日昌

（合肥工业大学 计算机与信息学院（人工智能学院），合肥 230601）

**摘要：**随着智能生成技术的不断发展与应用，手语数字人技术的应用场景也在不断扩大。智能生成技术可以将文本、语音等不同形式的信息转化为手语数字人的形式，为聋哑人士提供更加多样化和便捷的交流方式。此外，智能生成技术还可以通过机器学习和人工智能算法不断优化手语识别和生成的准确性与速度，提高手语数字人技术的用户体验和应用效果。本文将从人工智能的发展历程出发，详细介绍基于智能生成技术的手语数字人发展现状、所遇到的挑战，以及未来发展趋势。

**关键词：**智能生成技术；手语数字人；人工智能

**中图分类号：**TP37 **文献标志码：**A **文章编号：**2096-5036(2023)04-0020-12

**DOI：**10.16453/j.2096-5036.2023.04.003

## 0 引言

目前，人工智能技术不断发展，使得智能化系统和应用场景的范围不断扩大，为人们的生活和工作带来了极大的便利。然而，在这个快速变化的数字化时代，仍有一些人无法用语言进行有效的交流（如聋哑人士、失语症患者等）。手语数字人技术的出现，为这些人士提供了一种全新的交流方式，使他们能够用手语进行沟通和表达。通过手语数字人技术，聋哑人士可以用手语进行交互式的沟通和表达，促进社交互动、提高工作效率、辅助学习和享受娱乐体验等，在日常生活、工作场所、教育文化和社交媒体等领域都具有广泛的应用前景。同时，手语数字人技术还可以与其他技术相结合（如虚拟现实、增强现实等），进一步扩展其应用范围，为聋哑人士带来更加丰富的体验和更多的机遇。总之，手语数字人系统的开发和发展将有力地促进残障人士的社交融入和生活改善，也将为数字时代的发展和进步带来新的机遇和挑战。

## 1 人工智能发展介绍

人工智能（Artificial Intelligence，AI）自 20 世纪 50 年代初发展至今，已有近

基金项目：国家自然科学基金（U20A20183）；中央高校基本科研业务费专项资金（JZ2023HGQA0097）



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

70 年的发展历史，共历经三次发展高潮。第一次高潮始于 20 世纪 50 年代，早期的人工智能研究主要集中在推理和问题解决方面（如使用逻辑推理解决数学问题）。在 20 世纪 60 年代和 20 世纪 70 年代，人工智能的研究逐渐扩展到了更广泛的领域，包括机器翻译、语音识别、图像识别等。这些领域的研究成果为今后的人工智能技术奠定了基础，但是支撑研究的数据量并不充足，而且计算能力也十分有限。20 世纪 80 年代，神经网络算法和特定领域的专家系统等新技术的广泛应用迎来了人工智能发展的第二次高潮，这些技术使得人工智能的实现更加高效和智能化。同时，计算机的性能也得到了大幅提升，这为人工智能的发展提供了更好的硬件支持。进入 21 世纪，随着大数据和云计算等技术的发展，人工智能进入第三次高潮，人工智能已经涵盖了自然语言处理、计算机视觉、机器人技术、智能生成技术等领域，为人类的生产和生活带来了巨大的改变和便利。下文将具体对人工智能的三次发展高潮进行介绍。

20 世纪 50 年代，随着第一台通用计算机 ENIAC 的问世，打开了人工智能技术飞速发展的大门。1956 年，美国新罕布什尔州达特茅斯会议的成功举办标志着人工智能研究的起点，该会议旨在将计算机科学与认知科学相结合，研究机器可以如何模拟人类智能。在此期间，研究者们提出了一些早期的人工智能概念和算法，如逻辑推理和专家系统，例如由 Allen Newell 和 Herbert A. Simon 等编写的首个可以推理自动化的计算机程序 “Logic Theorist” [1]，它的关键思想是使用自动搜索和规则应用构建一个逻辑的证明树，可以通过不断地应用逻辑规则和公理，尝试从初始条件到达所需的结论，也可以自动选择和应用不同的规则，进行搜索和探索，并通过剪枝等技术提高搜索效率。然而，许多应用难题并没有随着时间推移而被解决，神经网络的研究也陷入停滞。

20 世纪 80 年代，迎来了人工智能发展的第二次高潮，神经网络和特定领域的专家系统等新的技术得到了广泛应用。BP 算法的提出，让非线性分类问题得以解决，并且具有较高的准确性和泛化能力。此外，特定领域的专家系统的兴起掀起了浪潮，代表性项目为 Edward Shortliffe 等开发的专家系统 “MYCIN” [2]。它的目标是利用人工智能技术辅助医生开展在感染性疾病的工作，展示了人工智能在医学领域的潜力，并为后来的医疗决策支持系统和临床决策辅助系统指明了发展方向。然而，BP 算法容易陷入局部最优解，需要较长的训练时间和大量的计算资源，特定专家系统也出现了数据获取困难，应用范围不广等问题。

从 2010 年开始，随着计算能力的提高和数据的大量积累，深度学习成为了当前人工智能研究的主流技术之一，掀起了人工智能第三次高潮。深度学习是一种通过多层的神经网络结构进行模式识别并特征提取的机器学习技术，目前在模式识别 [3]、图像生成 [4]、目标检测 [5] 等领域取得了很多突破性成果。深度学习的优势在于其具有强大的学习能力和表达能力，能够自动学习并提取数据中的特征，从而实现对大规模数据的高效处理和分析。这一技术的发展，为人工智能领域的研究和应用提供了全新的思路，尤其是智能生成技术，发展十分迅速，这让大模型也变得越来越流行。例如，2018 年谷歌发布的 BERT[6]（Bidirectional Encoder Representations from Transformers）是一种预训练的自然语言处理模型，采用了 Transformer 架构，并使用了双向编码器学



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

习句子的上下文信息，从而能够更好地理解自然语言中的语义和语法；2019 年，由卡内基梅隆大学（CMU）、Google Brain 和纽约大学等机构的研究人员提出的 XLNet[7] 是一种基于 Transformer 的预训练自然语言处理模型，采用了自回归和自编码两种预训练方式的结合，从而能够更好地理解自然语言中的语义和语法；2023 年，由 OpenAI 开发的 ChatGPT[8]（Chat Generative Pre-trained Transformer），一种基于 Transformer 架构的预训练语言模型，可以生成连贯和合理的自然语言文本，并具有很强的语义理解能力。它在多个 NLP 任务上取得了显著的成果，包括文本生成、机器翻译、文本摘要等。

随着时间的推移，人工智能领域的各个研究均取得了显著进展，但也面临一些挑战和问题。例如，数据隐私和伦理问题，人工智能对就业市场的影响，算法的公平性和透明性等都是当前人工智能发展中需要解决的重要议题。人工智能的发展是一个持续不断的过程，未来还将涌现出更多新的技术和应用。随着技术的不断进步和创新，人工智能将继续对我们的生活和社会产生更加深远影响。

## 2 基于深度学习的智能生成技术

智能生成技术是深度学习中的一个热门研究领域，涵盖了从早期的文本语音到后期的图像视频生成的技术。随着深度学习模型的不断发展，智能生成技术已经取得了很大的进展。在早期的文本语音生成 [9] 方面，深度学习模型主要应用于语音识别和自然语言处理。语音识别技术可以将语音信号转化为文本，使得机器能够理解人类的语言。自然语言处理技术则可以对文本进行分析和处理，实现自动化的文本生成、摘要和翻译等功能。随着深度学习模型的不断发展，智能生成技术开始应用于图像生成 [10] 方面。生成对抗网络是一种流行的深度学习模型，可以用于生成高质量的图像。近几年，智能生成技术还开始应用于视频生成 [11] 方面。深度学习模型可以通过学习视频序列中的特征生成新的视频内容（如视频剪辑和电影特效等）。另外，深度学习模型还可以用于视频超分辨率重建，提高视频的清晰度和细节。总之，随着深度学习技术的不断发展，智能生成技术在文本语音、图像和视频等方面都取得了很大进展。这些技术将会在各个领域产生广泛的应用，推动人工智能技术的发展和创新。

### 2.1 文本生成

循环神经网络（Recurrent Neural Networks，RNNs）（如图 1 所示）与 Transformer 是常用的文本生成模型，它们能够学习语言的语法和上下文关系，并生成具有一定连贯性和语义的文本。

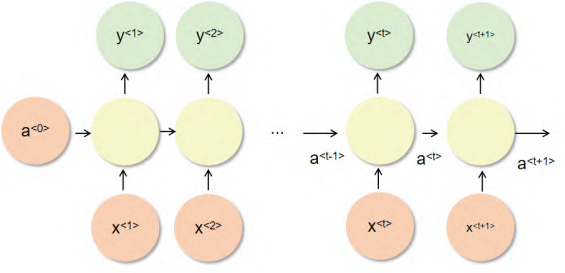


图 1 传统的循环神经网络（RNN）模型



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

RNN 是一种递归的神经网络结构，具有循环连接，使其能够处理具有时间依赖性的序列数据。循环神经网络通过将当前时刻的输入与前一时刻的隐藏状态进行组合，可以传递信息和记忆序列中的上下文信息。这使得循环神经网络在机器翻译、语音识别等方面有着广泛的应用。然而，传统的循环神经网络存在梯度消失或梯度爆炸等问题，使得其很难捕捉到长时间相依性。为了改善这个问题，出现了一些改进的 RNN 变体，如长短期记忆网络（Long Short-Term Memory，LSTM）和门控循环单元（Gated Recurrent Unit，GRU），通过引入门控机制，从而能够更好地控制信息的流动和记忆的更新。

Transformer[12] 是一种基于自注意力机制的序列建模模型，该模型最初应用于自然语言处理任务（如机器翻译等）。与传统的循环神经网络不同，Transformer 模型不需要循环连接，而是通过自注意力机制同时考虑序列中的所有位置。Transformer 模型通过编码器 - 解码器架构，将输入序列转化为中间表示，进而由解码器生成输出序列。自注意力机制能够在编码器与解码器之间建立全局关系，使得模型能够更好地捕捉长距离的依赖关系。这一创新性的模型，为序列建模任务的处理提供了全新思路，具有广泛的应用前景。

## 2.2 图像生成

以深度学习为基础的图像生成技术能够生成逼真的图像，包括自然景观、人脸、动物等。其中，生成对抗网络（Generative Adversarial Networks，GANs）是一种常见的方法（如图 2 所示），它由一个生成网络和一个判别网络组成，通过对抗学习的方式持续地对所产生的图像进行优化。生成器和判别器在对抗性中互相竞争，互相学习。在训练期间，生成器产生一组虚假样本，并把它们传送到判别器。判别器对这些样本进行分类，并返回分类结果。生成器根据判别器的反馈优化自己的生成策略，使生成的样本能够更好地欺骗判别器。同时，判别器也会根据生成器生成的样本更新自己的分类能力。生成对抗网络的优点是能够生成具有多样性和逼真度的样本数据，而无需显式地对生成过程进行建模。它在图像生成、图像修复、图像转换、语音合成等任务中取得了显著的成果。然而，GANs 也面临一些挑战，如训练的不稳定性、模式坍塌问题和生成样本的多样性控制等。因此，研究人员一直在努力改进生成对抗网络的模型结构，并推动其在各个领域的应用。

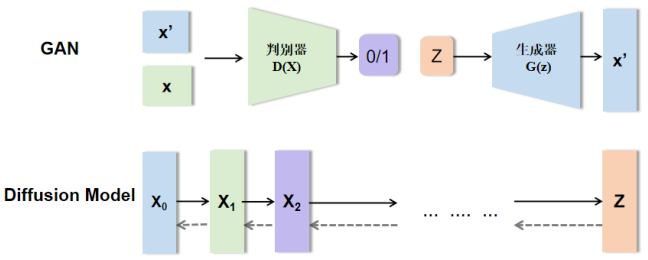


图 2 GAN 与 Diffusion Model 对比图

除了传统的生成对抗网络方法，近期比较火的图像生成模型还有 Diffusion Model[13]，一种基于去噪技术的图像生成模型（如图 3 所示）。Diffusion Model 的基本思想是通过一系列的逆向微分方程迭代地改变噪声信号，逐步逼近目标数据分布。



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

具体而言，Diffusion Model 将目标数据视为一个潜在噪声信号的转换过程，其中每个步骤都会引入一定的噪声，并通过一系列的逆向转换恢复出更接近目标数据的信号。这个过程可以看作是在随机性和确定性之间进行权衡的过程。

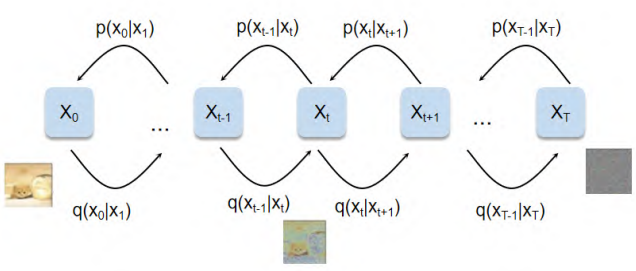


图 3 Diffusion Model 原理图

Diffusion Model 在图像生成、图像去噪、图像修复等任务中取得了显著的成果。它为建模复杂分布的数据提供了一种新的思路，并在机器学习和生成模型的研究中得到了广泛应用。

## 2.3 视频生成

针对视频生成，目前应用较广泛的是卷积神经网络（Convolutional Neural Networks，CNNs），可以提取视频的空间和时间特征，通过学习现有视频数据，生成具有连续动作和场景变化的新视频。为了处理视频数据，通常会使用 3D 卷积神经网络（3D Convolutional Neural Networks，3D CNNs）。3D 卷积神经网络通过在时间和空间维度上应用卷积操作提取时空特征。它在卷积层中同时考虑了图像的高度、宽度和时间维度，这样能够捕捉到视频中的动态特征。

近期，谷歌还发布了 Imagen Video，它采用了级联视频扩散模型，实现了基于文本条件的视频生成（如图 4 所示）。通过输入文本提示，该系统可以生成高清视频，其中包含了一个基础视频扩散模型、一个 frozen 文本编码器，以及一个空间和时间超分辨率模型。整个系统共计 116 亿个参数，在生成高质量视频方面具有很高的性能。此外，Imagen Video 还具有高度控制力和知识，可以产生多种不同的美术风格的影片、文字动画，并可以对三维物体进行理解。

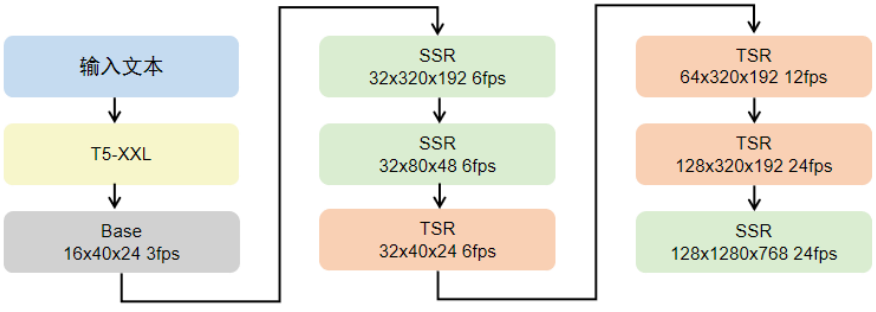


图 4 谷歌 Imagen Video 文本条件视频生成系统



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

## 3 手语数字人系统的发展现状

### 3.1 智能手语生成技术研究现状

在早期，手语动作生成主要侧重于对手语合成动画的研究，主要采用基于计算机图形学方面技术和统计模型等。在此基础上，通过收集不同的手语姿势样本，建立手语词汇姿势模型，再与手语库中的手语词组和单词相匹配，从而生成一段连贯、可视化的手语动画录像。例如，Glaubert 等 [14] 设计了一个语音手语助手交互系统，该系统使用语音或文本输入，并在特定主题领域内进行解析，将其转化为短语序列。然后，通过与手语数据库进行搜索匹配，系统找到与输入内容最匹配的手语短语。接下来，利用基于 SIGML 的手语动画转录技术，生成手语动画描述，最终输出完整的手语动画视频序列。还有一些研究为了给聋哑人士提供更多的教学资源，Karpouzis 及其团队 [15] 提出了一种方法，利用语法解析器分析书面文字的结构，并提取出关键信息和语义模式。然后，他们将这些结构模式与手语的对应模式进行匹配，以确定最佳的手语动作序列。通过这种方法，他们能够将书面文字转化为手语动画，实现文字和手语之间的转换。其中，标准虚拟字符动画技术被应用于合成手语动画序列，确保生成的手语动画具有准确性和流畅性。为了便于手语教育产业发展，Sagawa 等 [16] 专注于日本手语，开发了一种手语教学系统，其中包含手语识别和生成的功能。他们利用 3D 计算机图形动画技术，通过根据句子描述连接 3D 计算机图形的一系列参数，生成同步的手语动画。此外，该系统还具备自由改变手势方向和动画大小的功能，以增强用户的交互性和创造性。总而言之，以动画合成为基础的手语视频生成方法具有操作简便、效率高等优点，但是其性能在很大程度上依赖于构建大型的手语动画数据库。同时，合成的动画缺乏众多手势等逼真细节，对手语语义表达有限。因此，目前主流的研究趋势转向了更加逼真的手语智能生成技术研发。

随后，随着跨模态和图像生成技术的发展，更多的研究人员尝试将普通的手语动画生成转向了手语姿态视频的生成。Cui 等 [17] 提出了一种方法，将 DAE（Dropout Auto Encoder，DAE）与 LSTM 模型相结合。长短期记忆网络作为一种循环神经网络结构的变体，它可以通过最后一个动作，预测当前动作。而 DAE 通过对人体骨骼的隐性限制进行滤波，从而对手势动作进行最优化。Zelinka 等 [18] 提出了一种将前馈 Transformer 与循环 Transformer 相结合的优化姿势序列产生框架，用于提高手语姿势产生的效率与性能。除此之外，Xiao[19] 等还提出了一种以 VAE 为基础的概率骨架序列生成方法，其中利用基于 VAE 的编解码模型，生成具有随机性的姿态编码序列，并保持序列顺序和其他基本模式的不变性。Saunders 等 [20] 提出了一个对抗性多通道手语生成系统，它包括一个改进的 Transformer 生成器和条件判别器组成，它可以同时接受语音单词和手势，从而对手势序列的真伪做出判断。他们还通过端到端的方法将离散化的手势动作转换为连续化的表示，并在此基础上建立基于渐进式 Transformer 的手语生成模型。

除了手语姿态视频的生成，在目前阶段，我们还能够利用多阶段数据拟合产生逼真



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的手语视频。Ventura 等 [21] 则重点是探讨如何从 2D 姿态生成手语视频，并引入了一种独立的生成对抗网络，以捕捉脸部的细节，从而获得更加准确逼真的手语生成视频。而 Stoll 等 [22] 通过结合深度卷积生成对抗网络和卷积图像编码器的方法构建了一个手语视频生成系统。在此基础上，利用手势骨骼、表情等特征信息，通过图像编码器产生具有真实感的视频画面，并通过判别器对其进行评价。同时，将视频渲染技术引入到手语视频中，以增强其逼真度。

### 3.2 虚拟手语数字人技术发展动态

1985 年，哈拉维首次将虚拟数字人定义为有机体与无机物机器的结合体。在实际应用方面，虚拟数字人 [23] 的概念在 1982 年的动画作品《超时空要塞》中初次提出，引进了世界上第一位虚拟偶像“林明美”；随后，日本的 Crypton Future Media 在 2007 年发布了 VOCALOID 语音合成技术，创造出了虚拟偶像“初音未来”；2016 年，日本森仓圆设计的角色形象绊爱（Kizuna AI）在 YouTube 上线，成为了世界上第一个虚拟主播；2021 年，由清华大学计算机系、北京智源研究院、智谱 AI 和小冰公司联合培养的中国原创虚拟学生华智冰，通过对其进行不断的学习训练，展现了惊人的学习能力；京东云言犀团队 [24]，提出了一个多模态的话语决策模型，应用于客户服务中，包含了四个层次的知识体系，四十多个独立子系统，三千多个意图，三千万个问答知识点，涵盖了一千多万种自营商品的电商知识图谱，在为用户提供服务时不仅能解决用户需求，还能够考虑用户情绪、运用对话技术，提供可用、可控、可信的智能对话服务。除了语言，该平台还在画面、形象、仪态等方面实现了惟妙惟肖的呈现，通过人工智能大模型训练，充分提高了虚拟视觉的交互能力。可以看出，经过四十多年的发展历史，虚拟数字人技术的应用逐渐在娱乐、文化等多个领域活跃，但虚拟数字人的研究依旧面临一些挑战，如实现逼真的外观和动作、处理复杂的情感交互、提高自主决策能力和保护用户隐私等。然而，随着技术的不断进步和创新，虚拟数字人的研究前景仍然非常广阔，并且对于改善人机交互、创造沉浸式体验和提供个性化服务等方面具有巨大的潜力。

目前，随着虚拟视觉的火速发展，以及手语生成技术的逐渐完善，虚拟手语数字人技术开始出现在大众视野。在 2022 年冬奥会期间，由天津理工大学研发团队研发的央视新闻 AI 手语虚拟主播 [25,26] 的推出，首次让手语数字人进入到大众的视野中，在赛事期间让全国 2780 余万听障人士“听见”了北京冬奥会；近期，华为也发布了一款基于全属性特征识别与多模态基模融合的手语数字人，主要利用 3D 动画数字人的模型与 HMS Core 手语服务，产生手语老师、手语主播等不同的人物，其中包含了两万多个手语词汇，并支持 26 个脸部表情与精确的口型，在必要的情况下，可以进行适当的情绪表达，极大地提升了手语的可读性；百度智能云曦灵发布了一款“AI 手语平台”，该平台运用多种神经网络算法技术，研究将汉字转化为手语的算法，并构建出了一种基于神经网络的精确可控性手语翻译模型。该模型成功地将手语数字人翻译可读性提高到了超过 85% 的水平，并在手语合成视频的制作和现场手语主播直播方面得到了广泛应用，这一创新性的解决方案为我国手语服务的普及提供了全新的科学支持。



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

## 4 手语数字人技术的研究挑战

### 4.1 提高手语动作生成的效率和准确率

手语动作生成是将自然语言转化为手语动作的关键过程。在手语数字人技术中，为了提高手语动作生成的准确率，需要综合考虑手语的语种，手势的姿态和形态，手势的速度以及与自然语言的语义的对应关系。目前，手语动作的生成主要依赖于机器学习、深度学习等技术。通过训练模型，系统可以学习从自然语言到手语动作的映射关系，通过增加数据样本的种类和数量，进而提供更全面完善的训练数据，从而增强手语动作生成的准确率和模型泛化能力。利用旋转、缩放、平移等数据增强技术，得到更加丰富的姿态信息，从而提高识别精度；也可以通过改进模型结构和算法，进而提高手语动作生成的准确率。例如，采用更深层次、更复杂的神经网络结构，引入注意力机制或生成对抗网络等方法，提高模型的表达能力和生成效果。在生成过程中，也需要及时进行反馈和调整。通过引入实时反馈机制，可以根据生成的手语动作与目标语义之间的差异，对模型进行实时调整和优化，从而不断提升生成准确率。提高手语动作生成的准确率是手语数字人技术发展的重要目标之一，通过不断改进手语动作生成技术，实现更准确、流畅和自然的手语动作生成。

### 4.2 实现多样化的手语动作展示方式

手语动作展示是将生成的手语动作通过视觉形式进行可视化呈现的过程。在手语数字人技术中，手语动作的表达和展示需要注重其生动性和逼真性，以便听障人士能够更好地理解和学习手语。通过使用合适的手势动作、姿态和表情，手语数字人能够传达丰富的信息和情感，使得手语变得更加生动、易于理解和吸引人。目前，手语动作展示[27]主要采用手语合成动画、手语骨架视频、逼真手语视频等方式。随着技术的不断发展，手语数字人技术将结合多种虚拟视觉增强技术，为手语动作展示提供更多种可能性。例如，虚拟现实技术可以让听障人士身临其境地参与手语交流，增强其学习和理解手语的体验。增强现实技术可以将手语动作叠加在现实场景中，使得手语的展示更加直观和可感知。此外，三维重建技术和运动捕捉技术的应用也可以提高手语动作展示的逼真度和准确性。手语动作展示是手语数字人技术中至关重要的一环，它通过视觉化的方式使手语动作更具表达力、易于理解和吸引人。随着技术的不断创新和发展，手语数字人技术将为手语动作展示提供更多种可能性，为听障人士提供更好的手语学习和交流工具，并推动手语的传播和交流。

### 4.3 构建完整的手语数字人技术系统

构建一个完整全面的手语数字人技术体系将整合集成多个领域的技术，包括深度学习、自然语言处理、计算机视觉和虚拟视觉技术等。首先，需要对手语的基本语法，手势的形态、颜色和动态等特征进行详细的分析和研究，以确定手语的语义和规则。这涉及到对手势的编码和表示方法的设计，以及手语中手部动作和表情的语义解释。其次，



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

需要将这些手语的规则和特征与自然语言处理技术相融合。通过利用自然语言处理将手语转换为自然语言文本或语义表示，实现手语的自动识别和理解。通过深度学习技术，建立基于深度神经网络的手语识别与生成模型，并利用海量手语数据对其进行训练，从而提升模型的准确率与泛化性能。此外，将手语数字人技术与计算机图形学和图像处理技术相结合，可以实现手语动作更加精准和生动的呈现。通过使用计算机图形学技术，可以生成逼真的手语动画或虚拟手势，使其与手语识别和生成的结果相对应，并利用图像处理技术对手语图像和视频的预处理和特征提取，以增强手语识别和分析的效果。最后，要将多个领域的技术集成为一个完整的手语数字人体系，需要进行系统设计和整合。这涉及到设计高效的算法和模型，建立合适的数据集和测试环境，进行系统性能评估和优化。整个手语数字人体系的目标是实现手语的高效、自然和逼真的展示，使得手语成为一种更加普遍和无障碍的交流方式，实现对手语的全面理解、识别和生成，促进手语技术在实际应用中的发展和应用。

### 4.4 探索手语数字人技术的真实应用场景

目前，手语数字人技术还未被大范围在产业中应用，发展前景广阔。例如，在教育领域中，手语数字人技术可以应用于听障学生的手语教育中，提供可视化的手语学习工具和互动学习体验。通过与手语数字人的互动，学生可以学习手势、姿态和表情，提高手语沟通能力；在文化传播中，手语数字人技术可以在电影、电视剧、演唱会等文化活动中应用，为听障观众提供实时的手语翻译和展示。这能够使听障观众更好地理解和欣赏文化活动，促进文化的传播和包容性；在辅助通信中，手语数字人技术可以用于辅助听障人士的日常交流和沟通，通过手语数字人的实时翻译，听障人士可以正常与其他人进行交流，解决语言障碍问题；在虚拟现实和增强现实应用中，手语数字人技术可以与虚拟现实（VR）和增强现实（AR）相结合，为用户提供更加沉浸式的手语体验。用户可以通过佩戴 VR/AR 设备与手语数字人进行互动，进一步增强交互体验和学习效果。在开发和拓展手语数字人技术的真实应用场景时，需要考虑用户需求、技术可行性和实际可操作性。同时，也应与听障人士和相关领域的专业人士密切合作，获取他们的反馈和建议，以确保技术的质量和实用性，将手语数字人技术推向更广泛的应用场景。

## 5 基于智能生成技术的手语数字人未来发展展望

### 5.1 手语数字人的生成算法创新

手语数字人生成算法旨在利用自然语言处理、计算机视觉和计算机图形学等多个领域的技术，通过对语音或文本的处理，自动地生成与语义相匹配的手语动作，并以视觉形式进行展示。在手语数字人生成算法的研究中，面临着几个关键问题。首先，对于语音或文本的自动识别，需要开发智能算法准确地将语音或文本转化为机器可理解的形式。其次，对于手势的自动生成，需要将语义信息与手势的形态、姿态和速度等特征相结合，以生成自然流畅且符合语义的手语动作。最后，对于手语动作的实时展示，要求



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

算法能够以快速、稳定的方式将生成的手语动作呈现给用户，以实现实时交流和反馈。

为了更好地优化手语模型算法结构，传达更准确的语义信息，我们也可以构建手语知识图谱，完备手语相关知识体系的建立。知识图谱[28]是一种以图形形式展现的结构化数据集合，其中包含了丰富的实体以及实体之间的关系。构建手语实体的知识图谱具有重要意义，它可以显著提升手语识别、生成和翻译等领域的语义理解能力。为了构建这样的知识图谱，可以借助通用知识库，如 WordNet、ConceptNet，以及特定领域的知识库，将手语相关的知识元素与共性知识进行关联融合，从而构建出一个综合全面的手语知识图谱。

随着手语数字人生成算法的不断改进和完善，生成算法的准确性和流畅性将得到进一步提升，通过结合多领域技术和大量实验数据的探索，为听障人士提供更广泛、便捷的交流方式。

## 5.2 面向手语研究的真实数据扩充

拓展真实场景数据集是为了提供更真实、多样化的手语数据，以更好地应用于实际生活场景。然而，现有的手语数据大多来自于实验室，缺少了实际应用中的多样性和复杂性。因此，通过利用网络电视平台、社交媒体、在线视频等渠道，收集更多真实场景下的手语数据是必要的。例如，可以收集来自广播电视节目、新闻报道、社交媒体平台上的手语视频，涵盖各种日常生活情境和不同的手语表达方式。这样的真实场景数据集能够更好地反映手语在实际生活中的应用场景，从而提升手语识别、生成和翻译等技术的性能和适应性。

另外，在手语识别领域，由于目标域的手语训练数据往往不足，这成为了一个普遍存在的问题。然而，在其他领域，如新闻手语、气象手语等，往往会产生海量的跨领域数据。这为解决手语识别中的数据不足问题提供了新的机遇。通过利用这些跨域数据，可以扩充手语数据集，提供更丰富的训练样本。为了有效地利用跨域数据，需要克服源域数据和目标域数据之间的分布差异。如果这种差异得以消除，网络训练过程中数据不足对模型性能的影响将得到减轻，进而实现扩充手语数据集的目标。一种常见的方法是基于跨域知识迁移，通过识别不同领域数据之间的共性特征，以实现跨域数据的有效融合。另外，基于模型的迁移也是手语研究未来很有潜力的发展方向。通过将已经训练好的模型应用于目标域数据，我们可以借助预训练的知识改善目标任务的性能。因此，在手语识别领域，充分利用跨域数据和模型迁移的方法对于解决数据不足问题以及提升手语识别性能具有重要意义，这些方法的研究与探索将为手语研究领域的发展带来深远影响。除了利用多领域数据和模型迁移外，通过利用小样本或零样本学习的方法，对全新的词汇进行合理的语义推测，也可以解决手语识别中词汇增加的问题，并推动该领域的进一步发展。

## 5.3 手语数字人的应用体系构建

构建手语数字人体系是一个复杂而多层次的任务，它旨在创造一个能够准确、生动



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

地表达手语交流的数字化人机交互系统。这个体系涵盖了手语交流的各个层面，包括手势动作、表情、姿势，以及非手势元素等，以使得手语能够以多样化、丰富的方式进行表达。在构建手语数字人应用体系的过程中，关键是整合并利用大规模的手语数据集、手语知识图谱，以及计算机图形学和深度学习等技术手段。通过机器学习、姿态估计、动作识别和生成等技术，可以实现对手语动作的识别、生成和分析。这些技术能够从手势数据中提取关键信息，包括手势的形态、速度、动态特征等，并将其转化为可理解的数字化表示形式。同时，结合语义知识图谱和自然语言处理技术，可以实现对手语语义的理解和表达，从而使得手语数字人能够更好地传达意义和信息。此外，构建手语数字人体系有助于提升手语交流的便利性和可扩展性，同时也为手语教育、智能辅助手语翻译等领域提供更多可能性。在未来，更准确、更灵活的手语动作识别和生成技术将使手语数字人更加自然生动，能够更好地模拟人类手语交流，并且还将结合增强现实、虚拟现实等多种人机交互技术，进而提高手语数字人的交互性和真实感。

综上所述，通过构建手语数字人体系，我们可以更好地促进手语技术在实际生活中的落地和应用。这将为手语使用者提供更便捷、更自然的交流方式，推动手语识别、手语翻译等技术的不断创新和提升，以实现手语技术的全面发展和普及。

## 参考文献

[1] GUGERTY L. Newell and Simon's logic theorist: historical background and impact on cognitive modeling[C]//Proceedings of the human factors and ergonomics society annual meeting. Sage CA: Los Angeles, CA: SAGE Publications, 2006, 50(9): 880-884.
[2] EDWARD S. Computer-based medical consultations: MYCIN[M]. Elsevier, 1976.
[3] GHEYAS I A, SMITH L S. Feature subset selection in large dimensionality domains[J]. Pattern recognition, 2010, 43(1): 5-13.
[4] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.
[5] JIAO L, ZHANG F, LIU F, et al. A survey of deep learning-based object detection[J]. IEEE access, 2019, 7: 128837-128868.
[6] SUN F, LIU J, WU J, et al. BERT4Rec: sequential recommendation with bidirectional encoder representations from transformer[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. ACM, 2019: 1441-1450.
[7] YANG Z, DAI Z, YANG Y, et al. XLNet: generalized autoregressive pretraining for language understanding[J]. Advances in Neural Information Processing Systems, 2019, 32.
[8] SALLAM M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns[C]//Healthcare, 2023, 11(6): 887.
[9] NAKANO Y, SAEKI T, TAKAMICHI S, et al. vTTS: visual-text to speech[C]//2022 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2023: 936-942.
[10] ZHANG H, XU T, LI H, et al. StackGAN++: realistic image synthesis with stacked generative adversarial networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(8): 1947-1962.
[11] YANG Z, ZHU W, WU W, et al. Transmomo: invariance-driven unsupervised video motion retargeting[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 5306-5315.
[12] HAN K, XIAO A, WU E, et al. Transformer in transformer[J]. Advances in Neural Information Processing Systems, 2021, 34: 15908-15919.
[13] WIJMANS J G, BAKER R W. The solution-diffusion model: a review[J]. Journal of Membrane Science, 1995, 107(1-2): 1-21.
[14] GLAUERT J R W, ELLIOTT R, COX S J, et al. Vanessa-a system for communication between deaf and hearing people[J]. Technology and Disability, 2006, 18(4): 207-216.
[15] KARPOUZIS K, CARIDAKIS G, FOTINEA S E, et al. Educational resources and implementation of a greek sign language synthesis architecture[J]. Computers & Education, 2007, 49(1): 54-74.
[16] SAGAWA H, TAKEUCHI M. A teaching system of japanese sign language using sign language recognition and generation[C]//Proceedings of the Tenth ACM international conference on Multimedia. ACM, 2002: 137-145.



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[17] CUI R, CAO Z, PAN W, et al. Deep gesture video generation with learning on regions of interest[J]. IEEE Transactions on Multimedia, 2019, 22(10): 2551-2563.
[18] ZELINKA J, KANIS J. Neural sign language synthesis: words are our glosses[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 2020: 3395-3403.
[19] XIAO Q, QIN M, YIN Y. Skeleton-based Chinese sign language recognition and generation for bidirectional communication between deaf and hearing people[J]. Neural networks, 2020, 125: 41-55.
[20] SAUNDERS B, CAMGOZ N C, BOWDEN R. Adversarial training for multi-channel sign language production[J]. arXiv preprint arXiv:2008.12405, 2020.
[21] STOLL S, CAMGOZ N C, HADFIELD S, et al. Text2Sign: towards sign language production using neural machine translation and generative adversarial networks[J]. International Journal of Computer Vision, 2020, 128(4): 891-908.
[22] VENTURA L, DUARTE A, GIRÓ-I-NIETO X. Can everybody sign now? Exploring sign language video generation from 2D poses[J]. arXiv preprint arXiv:2012.10941, 2020.
[23] 郭全中. 虚拟数字人发展的现状、关键与未来[J]. 新闻与写作, 2022(7):56-64
[24] 张佳星. 多模态话语决策模型让机器人学会“捧哏”[N]. 科技日报,2023-05-09(06).
[25] 郑弘, 丰树琪. 人工智能在新闻报道中的突破、传播和价值——以总台央视新闻AI手语主播为例[J]. 传媒, 2022, (20):48-50.
[26] 陈望都, 林晓青. 为了听不到的你——世界杯首次数智手语解说技术详解[J]. 影视制作, 2023, 29(1): 26-32.
[27] 唐申庚. 基于深度学习的手语翻译与生成技术研究[D]. 合肥工业大学, 2022.
[28] 郭丹, 唐申庚, 洪日昌, 等.手语识别、翻译与生成综述[J]. 计算机科学, 2021, 48(3): 60-70.

**唐申庚**

合肥工业大学计算机与信息学院（人工智能学院）讲师。主要研究方向为机器学习与人工智能、多媒体解析与推理、手语视频翻译与生成。
* 通讯作者 email：tangsg@hfut.edu.cn

**修雪玉**

合肥工业大学计算机与信息学院（人工智能学院）本科生在读。主要研究方向为机器学习与模式识别、多媒体内容理解与分析。

**郭　丹**

合肥工业大学计算机与信息学院（人工智能学院）教授、博士生导师。主要研究方向为计算机视觉、机器学习与模式识别、智能多媒体内容分析。主持国家重点研发计划子课题、国家自然科学基金面上项目等基金项目 10 余项。

**洪日昌**

合肥工业大学计算机与信息学院（人工智能学院）教授、博士生导师。现任合肥工业大学计算机与信息学院常务副院长、安徽省人工智能学会理事长。研究成果获得 2015 年度国家自然科学二等奖、2017 年度安徽省自然科学一等奖和 2020 年度安徽省科技进步一等奖。曾获 2019 年安徽省青年五四奖章，入选中组部“万人计划”和教育部“长江学者奖励计划”。



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net