视频编辑、音频生成与扩散模型高质量论文与综述报告

1. 引言

生成式人工智能（AI）在多媒体内容创作领域展现出变革性的影响力。视频编辑和音频生成作为其中的关键组成部分，在从娱乐到专业内容制作的各种应用中日益重要。近年来，扩散模型作为一类强大的生成模型，彻底改变了这些领域。本报告旨在提供视频编辑、音频生成和扩散模型领域高质量研究论文和综述的专家级概述。

2. **深度学习在视频编辑中的应用**
    - 近期基于深度学习的视频编辑综述
      人工智能驱动的视频编辑旨在通过自动化任务和增强工作流程来提高效率，这一领域的研究兴趣日益浓厚 1。这些综述强调了对高效编辑解决方案不断增长的需求以及人工智能解决传统方法局限性的潜力。一篇综述 5 探讨了生成式人工智能和大型语言模型（LLM）在视频领域的交互，涵盖了 100 多篇论文，并注意到从文本到图像生成到文本到视频生成的快速发展。这篇综述强调了使用 LLM 进行视频创建、理解和交付的令人兴奋的可能性。另一篇综述 6 关注深度学习在视频分类中的创新，为与编辑相关的视频分析技术提供了背景。理解深度学习模型如何对视频内容进行分类是开发智能编辑工具的基础，例如自动分类场景或识别关键时刻。此外，一篇综述 7 专门探讨了视频实例分割的深度学习技术，这对于对象级视频编辑至关重要。精确识别和分割视频帧中的对象对于对象移除、替换或定向特效等任务至关重要。另一篇综述 8 强调了视频的表征学习，侧重于时空特征学习方法。有效的视频编辑通常依赖于理解单个帧的空间内容以及它们之间的时间关系，这使得表征学习成为一个关键领域。此外，一篇综述 9 重点介绍了视频分割的深度学习技术，这对于各种编辑操作至关重要。类似于实例分割，语义分割在实现内容感知的视频操作中也起着至关重要的作用。还有一篇综述 1 总结了自动视频编辑的发展历史以及人工智能在工作流程中的应用，涵盖了模态、输入类型、方法论、优化、数据集和评估指标。这篇全面的综述概述了该领域的发展以及人工智能视频编辑研究中考虑的各个方面。值得注意的是，还有专门针对基于扩散模型的视频编辑技术的综述 10，包括理论基础和应用 11。这突显了扩散模型作为视频处理强大工具的日益普及。这些大量涌现的综述表明该领域正在快速发展并日益复杂，研究人员需要全面的概述来了解最新的技术进展。视频编辑领域向扩散模型的转变是一个重要的趋势，因为这些模型在生成和操作任务中都展现出卓越的能力。
        - **洞察 1：** 综述的激增表明该领域正在快速发展并日益复杂，研究人员需要全面的概述来了解最新的技术进展。视频编辑领域向扩散模型的转变是一个重要的趋势。
    - 人工智能辅助视频编辑：方法与应用
      人工智能正被用于各种视频编辑任务，例如使用自然语言查找视频片段中的特定时刻、检测语气词、增强语音、推荐音乐以及添加补充镜头 12。Adobe 的

Project Blink 等项目展示了人工智能在简化常见编辑工作流程中的实际应用。大型语言模型（LLM）在视频理解方面展现出巨大的潜力，可以增强教育工具、用户界面和视频分析 5。LLM 能够提供对视频内容更高层次的理解，从而实现更智能、更用户友好的编辑功能。人工智能视频编辑器也已被开发用于实时事件和自动视频创建 13。 的工作展示了这方面的早期尝试，表明人们对自动化视频编辑过程的兴趣由来已久。与文本或基于图像的机器学习相比，视频机器学习面临着独特的挑战，包括需要复杂的运动和时间变化分析以及复杂的上下文理解 14。视频数据的时序维度以及视觉和听觉元素之间错综复杂的关系给机器学习模型带来了巨大的挑战。为了促进人工智能辅助视频组装领域的研究，超越视觉特效，已经开发了“视频编辑解剖学”等数据集 3。该数据集旨在推动人工智能视频编辑超越基本操作，朝着更具创造性和复杂性的任务发展。

- **洞察 2：** 人工智能正从基本操作转向需要理解视频内容和用户意图的更复杂任务。专业数据集的开发对于推进特定视频编辑应用的研究至关重要。

- 扩散模型在视频编辑中的应用

  专门针对基于扩散模型的视频编辑技术的综述 10 回顾了这些技术，包括理论基础和实际应用。这些综述全面概述了扩散模型（以其在图像生成方面的成功而闻名）如何在视频处理中得到应用。扩散模型视频编辑方法的分类是基于其底层技术 10 进行的。这种分类有助于理解这个快速发展领域中不同的策略和创新。V2VBench 等基准的引入是为了评估文本引导的视频编辑任务 10。基准对于客观比较不同的方法和跟踪领域的进展至关重要。16 中提到了基于扩散的视频生成在运动一致性和计算效率方面面临的挑战。虽然扩散模型在质量方面表现出色，但确保视频中的时间连贯性并降低其高计算成本是关键的研究挑战。SAVE 和 LatentEditor 等高效扩散模型在视频编辑中的应用旨在最大限度地减少资源需求，同时保持高质量 17。这些工作展示了使扩散模型更适用于实际视频编辑场景的努力。对视频扩散模型的综述 18 将工作分为视频生成、编辑和理解三类。这篇综述提供了关于扩散模型如何影响整个视频领域的更广泛的视角。对视频生成扩散模型的综述 19 全面概述了扩散模型，包括应用、架构和时间动态。虽然侧重于生成，但许多这些技术可以转移到编辑。Awesome-Video-Diffusion-Models GitHub 存储库 20 等资源提供了一个精选的论文和模型列表。此类资源对于研究人员来说非常宝贵，可以让他们及时了解最新的进展。

  - **洞察 3：** 扩散模型正因其生成高质量和时间连贯结果的能力而成为视频编辑领域的主流方法。研究正积极解决计算成本问题，并探索各种条件方法以实现可控编辑。

- 挑战与未来方向

  人工智能视频编辑仍面临着一些开放的挑战，例如多模态集成、理解用户意图以及确保逻辑和美学上的连贯性 1。视频通常包含音频、文本和复杂的叙述，需要人工智能以连贯的方式处理和理解这些不同的元素。未来的研究方向包括提高自动化程度、增强视频理解以及开发更智能的编辑工具 1。最终目标是创建能够以更有意义和更直观的方式协助视频编辑的人工智能系统，甚至可以自动化编辑过

程中的大部分工作。

3. **神经音频合成与生成**
    - 神经音频合成的里程碑论文
      WaveNet 21 是一项开创性的工作，它使用卷积神经网络（CNN）进行原始音频生成，显著提高了语音合成的质量。WaveNet 直接对原始音频波形进行建模的能力彻底改变了该领域，并为随后的发展奠定了基础。22 中讨论的“使用 WaveNet 自编码器进行音乐音符的神经音频合成”介绍了 NSynth 数据集和一个 WaveNet 风格的自编码器，用于音乐音符合成和音色插值。这项工作展示了神经网络学习和操作乐器声音细微差别的潜力。GANSynth 25 通过对数幅度建模和瞬时频率，使用生成对抗网络（GAN）展示了高保真和快速的音频生成。GANSynth 提供了一种比 WaveNet 等自回归模型更有效的方法，同时在某些方面实现了相当甚至更好的质量。DDSP（可微分数字信号处理）27 将经典的信号处理与深度学习相结合，用于程序化音频合成。DDSP 允许对音频参数进行可解释的控制，弥合了传统合成技术和深度学习之间的差距。
        - **洞察 4：** 这些里程碑式的论文代表了神经音频合成发展中的关键节点，从像 WaveNet 这样的自回归模型发展到基于 GAN 的方法以提高效率和控制，并探索像 DDSP 这样的混合模型。
    - 神经音频合成综述
      29 中关于语音和视觉系统中深度神经网络的综述为神经音频合成提供了更广泛的背景。这篇综述强调了深度学习在不同感官模式中的总体影响。30 中关于音频合成和视听多模态处理的综述涵盖了文本到语音（TTS）、音乐生成和视听任务。这篇综述强调了多模态方法在音频研究中日益增长的重要性。31 中专门关于神经语音合成的综述涵盖了文本分析、声学模型和声码器。这提供了关于使用文本生成语音的技术的重点概述。34 中关于深度学习音频生成方法的综述对音频表示、架构（包括扩散模型）和评估指标进行了分类。这篇综述全面介绍了用于生成不同类型音频的各种深度学习技术。35 中关于音乐神经音频合成的综述，特别是快速推理模型。加速音频合成对于实时应用和交互式系统至关重要。
        - **洞察 5：** 这些综述的出现有助于研究人员理解神经音频合成的不同方面，从语音到音乐，以及所采用的各种深度学习技术。这些综述中越来越多地包含扩散模型表明了它们日益增长的重要性。
    - 扩散模型在音频生成中的应用
      DiffWave 36 是一种扩散概率模型，用于高保真条件和无条件波形生成。
      DiffWave 表明，扩散模型可以在音频合成中实现最先进的结果，并具有快速的推理时间。38 中专门关于用于 TTS 和语音增强的音频扩散模型的综述。这篇综述重点介绍了如何使用扩散模型来解决语音处理中的关键挑战。42 中提到了在具有全局和局部基于文本的条件下的文本到音乐生成中扩散模型的使用。这展示了扩散模型在文本描述引导下生成复杂音乐结构的能力。AudioLDM 43 是一种用于文本到音频生成的潜在扩散模型。在潜在空间中操作可以更有效地训练和生成高质量音频。扩散模型在视频到音频合成中的应用 44。这个新兴领域旨在为无声

视频生成同步音频，从而增强生成视频内容的真实感。34 中回顾深度学习音频生成方法（包括扩散模型）的综述。这篇综述为理解扩散模型在音频生成技术更大范围内的作用提供了更广阔的背景。46 中强调了在 ICASSP 和 INTERSPEECH 56 上展示的近期使用扩散模型进行音频生成的工作。扩散模型在各种音频生成任务中的普及，以及相关的综述和会议报告，突显了其有效性。对不同条件方法（例如，文本、视频）和架构的探索表明，人们正在不断努力充分利用扩散模型在音频领域的潜力。

- **洞察 6：** 扩散模型已成为各种音频生成任务的强大工具，在某些领域实现了高保真度并超越了先前的最先进方法。研究正积极探索扩散框架内不同的条件技术和架构。

- 顶级会议上音频生成的最新进展

  总结了 ICASSP 2024 55 和 2025 46 上提出的音频生成方面的关键论文和趋势。这些包括使用 LLM 改进语音识别、人工智能和信号处理的融合以及生成式数据增强方面的进展。讨论了 INTERSPEECH 2024 67 和 2025 56 的著名工作。关键主题包括多语种语音处理、视听语音识别和音频模型的可解释性。重点介绍了 IEEE/ACM Transactions on Audio, Speech, and Language Processing (T-ASLP) 74 和 The Journal of the Acoustical Society of America (JASA) 80 上发表的与音频生成扩散模型相关的论文。这些出版物代表了该领域的高影响力研究。

  - **洞察 7：** 信号处理和语音/音频领域的顶级会议正积极推出关于音频生成扩散模型的研究，这表明了社区的浓厚兴趣和该领域的快速进展。

4. **扩散模型：基础与应用**
   - 基本概念与数学公式

     扩散模型的核心思想是学习逆转噪声添加过程以生成数据 84。这涉及一个逐渐添加噪声的前向（扩散）过程和一个学习从噪声中生成数据的反向（去噪）过程 84。前向过程将数据转换为简单的噪声分布，而反向过程学习将此噪声映射回数据分布。关键的数学概念包括马尔可夫链、高斯噪声以及用于训练的变分下界（ELBO）84。这些概念为扩散模型框架提供了理论基础。提到了不同的公式，例如去噪扩散概率模型（DDPM）和基于分数的生成模型 84。DDPM 直接预测要去除的噪声，而基于分数的模型估计数据分布的梯度。

     - **洞察 8：** 理解潜在的数学原理对于理解扩散模型的能力和局限性至关重要。与热力学的联系为生成过程提供了独特的视角。

   - 有影响力的扩散模型架构与训练技术

     讨论了通常使用（通常经过修改的）U-Net 架构作为扩散模型的常见骨干网络，尤其是在图像和视频应用中 86。U-Net 的编码器-解码器结构与跳跃连接非常适合捕获全局上下文和细粒度细节。强调了扩散Transformer（DiT）作为一种利用 Transformer网络进行扩散建模的新颖架构的出现 91。DiT 在图像生成任务中展现出了卓越的可扩展性和性能。回顾了先进的训练技术，例如噪声调度、方差学习、无分类器引导和高效采样方法 84。这些技术对于优化训练过程、控制生成和提高生成样本的质量至关重要。提到了提高训练效率的技术，例如多阶段框架

和解决梯度差异 95。这些努力旨在降低训练扩散模型所需的计算成本和时间。讨论了使用分类器引导和提示工程等技术来控制生成的方法 98。引导扩散过程可以生成符合特定条件或用户指令的样本。

- **洞察 9：** 扩散模型的架构正在不断发展，Transformer 在已建立的 U-Net 之外展现出令人期待的结果。先进的训练技术对于实现高质量的结果和高效的训练至关重要。

- 扩散模型近期综述
  回顾了 103 中涵盖方法、应用、高效采样和改进似然估计的全面综述。讨论了 107 中专注于跨不同领域的异常检测的扩散模型综述。重点介绍了 109 中专门针对视频扩散模型的综述，涵盖了基础、实现和应用。提到了 38 中关于音频扩散模型在 TTS 和语音增强方面的综述。
  - **洞察 10：** 这些综述为理解扩散模型在不同数据模态和应用领域的广度和深度研究提供了宝贵的资源。
- 扩散模型的新颖应用
  提到了扩散模型在建筑设计中的应用 111。这展示了扩散模型在创意设计过程中的潜力。讨论了它们在蛋白质科学和医疗保健中的使用 114。这突显了扩散模型在科学发现中的适用性。重点介绍了在生成 3D 内容方面的应用 115。这展示了扩散模型处理超越 2D 图像的复杂数据结构的能力。提到了它们在天气预报和气候预测中的潜力 116。这表明扩散模型可用于建模复杂的物理系统。探索了它们在机器人操作中的使用 117。这展示了扩散模型在机器人和控制领域的应用。
  - **洞察 11：** 扩散模型正被证明是多功能的工具，其应用范围已超出传统的图像和视频生成，扩展到各种科学和创意领域。

5. **结论**
   深度学习在视频编辑、神经音频合成与生成以及扩散模型领域都取得了显著的进步。人工智能驱动的视频编辑正在利用深度学习来自动化复杂任务，增强用户体验，并为创意表达开辟新的途径。神经音频合成已经从开创性的 WaveNet 等模型发展到更高效且可控的 GAN 和混合方法，扩散模型在文本到语音、音乐生成和视频到音频合成等任务中展现出了卓越的能力。扩散模型本身已成为一种强大的生成建模技术，其应用范围已超越传统的图像和视频生成，扩展到建筑设计、蛋白质科学、天气预报和机器人等领域。这些领域日益融合，尤其是在视频和音频任务中越来越多地使用扩散模型，预示着多媒体内容创作和分析的未来将更加令人兴奋。这项研究的跨学科性质，借鉴了计算机视觉、自然语言处理、音频处理甚至物理学等多个领域的知识，进一步凸显了其重要性和潜力。随着这些技术的不断发展，它们有望彻底改变我们创作、编辑和体验多媒体内容的方式。

关键表格

1. **表 1：近期综述总结**

综述标题	重点领域	涵盖的关键主题	引用（片段 ID）
AI 视频编辑：综述	视频编辑	自动视频编辑的发展历史、人工智能在工作流程中的应用、模态、输入视频类型、方法论、优化、数据集、评估指标	1
生成式人工智能和 LLM 在视频生成、理解和流媒体方面的综述	视频编辑	生成式人工智能和 LLM 与视频领域的交互、视频生成、视频理解、视频流媒体	5
视频分类中深度学习的创新：技术和数据集评估综述	视频编辑	视频分类技术、数据集、网络架构、模型评估指标、并行处理方法	6
视频实例分割的深度学习技术：综述	视频编辑	视频实例分割、架构范式、功能性能比较、模型复杂度、计算开销	7
视频表征学习：综述	视频编辑	时空特征学习方法、空间特征、时间特征、挑战	8
视频分割深度学习技术综述	视频编辑	通用对象分割、视频语义分割、任务设置、背景概念、发展历史、主要挑战	9
基于扩散模型的视频编辑：综述	视频编辑	扩散模型、视频编辑技术、理论基础、实际应用、基准测试	10
深度神经网络语音和视觉系统综述	音频生成	最先进的深度神经网络架构、算法和系统在视觉和语音应用中的应用	29

音频合成与视听多模态处理综述	音频生成	音频合成、视听多模态处理、文本到语音、音乐生成、视听信息结合的任务	30
神经语音合成综述	音频生成	神经语音合成、文本分析、声学模型、声码器、快速 TTS、低资源 TTS、鲁棒 TTS、表现力 TTS、自适应 TTS	31
深度学习音频生成方法综述	音频生成	音频表示、深度学习架构（包括扩散模型）、评估指标	34
神经音频合成综述	音频生成	神经音频合成、音乐合成、快速推理模型	35
音频扩散模型综述：生成式人工智能中的文本到语音合成和增强	音频生成	音频扩散模型、文本到语音合成、语音增强	38
深度学习音频生成方法综述	音频生成	音频表示、深度学习架构、评估指标（包括扩散模型）	34
扩散模型：方法与应用综合综述	扩散模型	高效采样、改进的似然估计、处理特殊结构的数据、应用领域	103
扩散模型异常检测综述	扩散模型	扩散模型、异常检测、网络安全、欺诈检测、医疗保健、制造业	107
视频扩散模型综述：基础、实现与应用	扩散模型	扩散模型、视频生成、时间一致性、视觉质量、架构创新、优化策略	109

2. **表 2：神经音频合成的里程碑论文**

论文标题	主要贡献	使用的架构	引用（片段 ID）
WaveNet：原始音频的生成模型	使用卷积神经网络生成原始音频，显著提高语音合成质量	卷积神经网络	21
使用 WaveNet 自编码器进行音乐音符的神经音频合成	介绍了 NSynth 数据集，使用 WaveNet 风格的自编码器进行音乐音符合成和音色插值	WaveNet 风格的自编码器	22
GANSynth：对抗性神经音频合成	使用 GAN 通过建模对数幅度和瞬时频率生成高保真和快速的音频	生成对抗网络 (GAN)	25
DDSP：可微分数字信号处理	集成经典信号处理与深度学习，用于程序化音频合成	可微分数字信号处理	27

3. **表 3：扩散模型的代表性应用**

应用领域	具体任务	引用（片段 ID）
图像生成	高分辨率图像合成、文本到图像生成、图像修复	103
视频生成与编辑	文本到视频生成、图像到视频生成、视频修复、视频编辑	106
音频生成	文本到语音合成、文本到音乐生成、音频修复、语音增强、视频到音频合成	34
异常检测	网络安全、欺诈检测、医疗保健、制造业中的异常检测	107
建筑设计	生成建筑平面图、生成建筑模	111

	型	
蛋白质科学与医疗保健	蛋白质结构预测、药物发现	114
3D 内容生成	生成 3D 模型、神经辐射场 (NeRF)	113
天气预报与气候预测	模拟大气和海洋动力学、生成天气预报集成	116
机器人操作	轨迹规划、抓取预测	117

Works cited

1. (PDF) AI Video Editing: a Survey - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/357587491_AI_Video_Editing_a_Survey
2. AI Video Editing: a Survey[v2] - Preprints.org, accessed April 28, 2025, https://www.preprints.org/manuscript/202201.0016/v2
3. AI Video Editing: a Survey - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/358434018_AI_Video_Editing_a_Survey
4. A Survey on Content-Aware Image and Video Retargeting | Request PDF - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/326594846_A_Survey_on_Content-Aware_Image_and_Video_Retargeting
5. A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming, accessed April 28, 2025, https://arxiv.org/html/2404.16038v1
6. (PDF) Deep Learning Innovations in Video Classification: A Survey on Techniques and Dataset Evaluations - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/382196647_Deep_Learning_Innovations_in_Video_Classification_A_Survey_on_Techniques_and_Dataset_Evaluations
7. [2310.12393] Deep Learning Techniques for Video Instance Segmentation: A Survey - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2310.12393
8. Deep Video Representation Learning: a Survey - arXiv, accessed April 28, 2025, https://arxiv.org/html/2405.06574v1
9. A Survey on Deep Learning Technique for Video Segmentation - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/353053599_A_Survey_on_Deep_Learning_Technique_for_Video_Segmentation
10. Diffusion Model-Based Video Editing: A Survey - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/382145911_Diffusion_Model-Based_Video_Editing_A_Survey
11. [2407.07111] Diffusion Model-Based Video Editing: A Survey - arXiv, accessed

April 28, 2025, https://arxiv.org/abs/2407.07111
12. Project Blink: Creating the Future of AI-Powered Video Editing - Adobe Research, accessed April 28, 2025, https://research.adobe.com/news/project-blink-creating-the-future-of-ai-powered-video-editing/
13. AI video editing tools, accessed April 28, 2025, https://www.hsu-hh.de/imb/wp-content/uploads/sites/677/2021/08/AI-video-editing-tools-What-editors-want-and-how-far-is-AI-from-delivering.pdf
14. How much harder is it to do ML for video compared to text/images? [D] : r/MachineLearning, accessed April 28, 2025, https://www.reddit.com/r/MachineLearning/comments/18hs1vg/how_much_harder_is_it_to_do_ml_for_video_compared/
15. Innovative Deep Learning-based Video Editing Tool | Request PDF - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/358958653_Innovative_Deep_Learning-based_Video_Editing_Tool
16. Survey of Video Diffusion Models: Foundations, Implementations, and Applications, accessed April 28, 2025, https://openreview.net/forum?id=2ODDBObKjH
17. Effective and Efficient Use of Diffusion Models for Editing in Computer Vision - ucf stars, accessed April 28, 2025, https://stars.library.ucf.edu/etd2024/37/
18. [2310.10647] A Survey on Video Diffusion Models - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2310.10647
19. Video Diffusion Models: A Survey - OpenReview, accessed April 28, 2025, https://openreview.net/forum?id=rJSHjhEYJx
20. ChenHsing/Awesome-Video-Diffusion-Models: [CSUR] A ... - GitHub, accessed April 28, 2025, https://github.com/ChenHsing/Awesome-Video-Diffusion-Models
21. A Short History of Neural Synthesis - The Royal Northern College of Music, accessed April 28, 2025, https://www.rncm.ac.uk/research/research-activity/research-centres-rncm/prism/prism-blog/a-short-history-of-neural-synthesis/
22. [1704.01279] Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders - arXiv, accessed April 28, 2025, https://arxiv.org/abs/1704.01279
23. Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders - Proceedings of Machine Learning Research, accessed April 28, 2025, http://proceedings.mlr.press/v70/engel17a/engel17a.pdf
24. arxiv.org, accessed April 28, 2025, https://arxiv.org/pdf/1704.01279
25. GANSynth: Adversarial Neural Audio Synthesis - OpenReview, accessed April 28, 2025, https://openreview.net/forum?id=H1xQVn09FX
26. [R] GANSynth: Adversarial Neural Audio Synthesis (ICLR 2019) - Reddit, accessed April 28, 2025, https://www.reddit.com/r/MachineLearning/comments/avxddd/r_gansynth_adversarial_neural_audio_synthesis/
27. Procedural Engine Sounds Using Neural Audio Synthesis - DiVA portal, accessed April 28, 2025,

http://www.diva-portal.org/smash/get/diva2:1465597/FULLTEXT01.pdf
28. Audio Generation | Papers With Code, accessed April 28, 2025, https://paperswithcode.com/task/audio-generation
29. Survey on Deep Neural Networks in Speech and Vision Systems - PMC, accessed April 28, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7584105/
30. (PDF) A Survey on Audio Synthesis and Audio-Visual Multimodal Processing, accessed April 28, 2025, https://www.researchgate.net/publication/353654033_A_Survey_on_Audio_Synthesis_and_Audio-Visual_Multimodal_Processing
31. [2106.15561] A Survey on Neural Speech Synthesis - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2106.15561
32. tts-tutorial/survey: A Survey on Neural Speech Synthesis https://arxiv.org/pdf/2106.15561.pdf - GitHub, accessed April 28, 2025, https://github.com/tts-tutorial/survey
33. arxiv.org, accessed April 28, 2025, https://arxiv.org/pdf/2106.15561
34. A survey of deep learning audio generation methods - arXiv, accessed April 28, 2025, https://arxiv.org/pdf/2406.00146
35. Accelerating Neural Audio Synthesis - Research Collection, accessed April 28, 2025, https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/571861/2/Volhejn_Vaclav.pdf
36. diffwave:aversatile diffusion model for audio synthesis, accessed April 28, 2025, https://www.audiolabs-erlangen.de/content/05_fau/professor/00_mueller/02_teaching/2024s_sarntal/02_group_SYNTH/2022_Kong_DiffWave_arxiv.pdf
37. DiffWave: A Versatile Diffusion Model for Audio Synthesis - OpenReview, accessed April 28, 2025, https://openreview.net/forum?id=a-xFK8Ymz5J
38. A Survey on Audio Diffusion Models: Text To Speech Synthesis and Enhancement in Generative AI (2023) | Chenshuang Zhang | 35 Citations - SciSpace, accessed April 28, 2025, https://scispace.com/papers/a-survey-on-audio-diffusion-models-text-to-speech-synthesis-2so77w2i
39. arxiv.org, accessed April 28, 2025, https://arxiv.org/abs/2303.13336
40. (PDF) A Survey on Audio Diffusion Models: Text To Speech Synthesis and Enhancement in Generative AI - ResearchGate, accessed April 28, 2025, https://www.researchgate.net/publication/369477230_A_Survey_on_Audio_Diffusion_Models_Text_To_Speech_Synthesis_and_Enhancement_in_Generative_AI
41. arxiv.org, accessed April 28, 2025, https://arxiv.org/pdf/2303.13336
42. Diffusion based Text-to-Music Generation with Global and Local Text based Conditioning, accessed April 28, 2025, https://research.samsung.com/blog/Diffusion-based-Text-to-Music-Generation-with-Global-and-Local-Text-based-Conditioning
43. Sound Scene Synthesis at the DCASE 2024 Challenge - arXiv, accessed April 28, 2025, https://arxiv.org/html/2501.08587v1
44. DIFF-FOLEY: Synchronized Video-to-Audio Synthesis with Latent Diffusion Models, accessed April 28, 2025,

https://proceedings.neurips.cc/paper_files/paper/2023/file/98c50f47a37f63477c01558600dd225a-Paper-Conference.pdf
45. [2406.00146] A Survey of Deep Learning Audio Generation Methods - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2406.00146
46. ICASSP 2025 – Advances in speech recognition + gen AI integration - LXT, accessed April 28, 2025, https://www.lxt.ai/blog/icassp-2025-advances-in-speech-recognition-gen-ai-integration/
47. ICASSP - Amazon Science, accessed April 28, 2025, https://www.amazon.science/tag/icassp
48. Sony Shares New Research at International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025), accessed April 28, 2025, https://www.sony.com/en/SonyInfo/technology/stories/entries/ICASSP2025_report/
49. Audio and Speech Processing Jan 2025 - arXiv, accessed April 28, 2025, http://www.arxiv.org/list/eess.AS/2025-01?skip=0&show=250
50. GenDA 2025 - Google Sites, accessed April 28, 2025, https://sites.google.com/view/genda2025
51. List of Accepted Papers - IEEE ICASSP 2020 || Barcelona, Spain || 4-8 May 2020, accessed April 28, 2025, https://cmsworkshops.com/ICASSP2020/Papers/AcceptedPapers.asp
52. Building technologies to expand the future of sound for creators - Sony, accessed April 28, 2025, https://www.sony.com/en/SonyInfo/technology/stories/entries/interview_de_mitsufuji/
53. ICASSP: What “signal processing” has come to mean - Amazon Science, accessed April 28, 2025, https://www.amazon.science/blog/icassp-what-signal-processing-has-come-to-mean
54. icassp 2025, accessed April 28, 2025, https://2025.ieeeicassp.org/
55. aasp-l3.5: generation or replication: auscultating audio latent diffusion models - IEEE ICASSP 2024 || Seoul, Korea || 14-19 April 2024, accessed April 28, 2025, https://cmsworkshops.com/ICASSP2024/view_paper.php?PaperNum=7649
56. Challenges - Interspeech 2025, accessed April 28, 2025, https://www.interspeech2025.org/challenges
57. Accepted Tutorials - Interspeech 2025, accessed April 28, 2025, https://www.interspeech2025.org/tutorials?version=latest
58. Interspeech - Amazon Science, accessed April 28, 2025, https://www.amazon.science/tag/interspeech
59. Unconditional Audio Generation with Generative Adversarial Networks and Cycle Regularization - ISCA Archive, accessed April 28, 2025, https://www.isca-archive.org/interspeech_2020/liu20i_interspeech.html
60. Interspeech 2020 - ISCA Archive, accessed April 28, 2025, https://www.isca-archive.org/interspeech_2020/
61. Special Session on Interpretability in Audio and Speech Technology Interspeech

2025 - Google Sites, accessed April 28, 2025, https://sites.google.com/view/interspeech2025-interpret/home
62. INTERSPEECH: Conference of the International Speech Communication Association - WikiCFP, accessed April 28, 2025, http://www.wikicfp.com/cfp/program?id=1624
63. Takeaways from Interspeech 2020, Part 2: scientific highlights | Sonos Tech Blog, accessed April 28, 2025, https://tech-blog.sonos.com/posts/takeaways-from-interspeech-2020-part-2/
64. AccentBox: Towards High-Fidelity Zero-Shot Accent Generation - arXiv, accessed April 28, 2025, https://www.arxiv.org/pdf/2409.09098
65. [Blog] Speaker Personalization for Automatic Speech Recognition using Weight-Decomposed Low-Rank Adaptation - Samsung Research, accessed April 28, 2025, https://research.samsung.com/blog/Speaker-Personalization-for-Automatic-Speech-Recognition-using-Weight-Decomposed-Low-Rank-Adaptation
66. [2407.10468] LiteFocus: Accelerated Diffusion Inference for Long Audio Synthesis - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2407.10468
67. INTERSPEECH.2024 - Speech Synthesis | Cool Papers - Immersive Paper Discovery, accessed April 28, 2025, https://papers.cool/venue/INTERSPEECH.2024?group=Speech%20Synthesis
68. Interspeech 2024 - ISCA Archive, accessed April 28, 2025, https://www.isca-archive.org/interspeech_2024/index.html
69. INTERSPEECH.2024 - Speech Detection | Cool Papers - Immersive Paper Discovery, accessed April 28, 2025, https://papers.cool/venue/INTERSPEECH.2024?group=Speech%20Detection
70. DiffATR: Diffusion-based Generative Modeling for Audio-Text Retrieval - ISCA Archive, accessed April 28, 2025, https://www.isca-archive.org/interspeech_2024/xin24_interspeech.pdf
71. The X-LANCE Technical Report for Interspeech 2024 Speech Processing Using Discrete Speech Unit Challenge - arXiv, accessed April 28, 2025, https://arxiv.org/html/2404.06079v2
72. Provisional Programme | Interspeech 2024, accessed April 28, 2025, https://interspeech2024.org/wp-content/uploads/IS24-Provisional-Programme.pdf
73. ICASSP 2024: Seoul, Korea - DBLP, accessed April 28, 2025, https://dblp.org/db/conf/icassp/icassp2024
74. IEEE Transactions on Audio, Speech and Language Processing ..., accessed April 28, 2025, https://signalprocessingsociety.org/publications-resources/ieee-transactions-audio-speech-and-language-processing
75. TASLPRO Volume 33 | 2025 | IEEE Signal Processing Society, accessed April 28, 2025, https://signalprocessingsociety.org/publications-resources/ieee-transactions-audio-speech-and-language-processing/2025/01
76. IEEE Transactions on Audio, Speech & Language Processing, Volume 20 - DBLP,

accessed April 28, 2025, https://dblp.org/db/journals/taslp/taslp20
77. ISCApad #297 - ISCA Services, accessed April 28, 2025, https://services.isca-speech.org/iscapad/iscapad.php?module=category&id=2214
78. CfP IEEE JSTSP Special Issue on Tensor Decomposition for Signal Processing and Machine Learning - ISCA, accessed April 28, 2025, https://services.isca-speech.org/iscapad/iscapad.php?module=category&id=1836
79. IEEE/ACM transactions on audio, speech, and language processing - SciSpace, accessed April 28, 2025, https://scispace.com/journals/ieee-acm-transactions-on-audio-speech-and-language-1tsuthgq
80. Call for papers | The Journal of the Acoustical Society of America ..., accessed April 28, 2025, https://pubs.aip.org/asa/jasa/pages/specialissues
81. Special Issue: Generative and Physics-Informed Artificial Intelligence for Acoustics, accessed April 28, 2025, https://pubs.aip.org/asa/jasa/pages/cfp030425
82. ASA Publications - The Acoustical Society of America, accessed April 28, 2025, https://acousticalsociety.org/asa-publications/
83. Papers Wanted for JASA Special Issue on Statistical Science in AI | Amstat News, accessed April 28, 2025, https://magazine.amstat.org/blog/2024/04/01/papers-wanted-for-jasa-ai/
84. Introduction to Diffusion Models for Machine Learning - AssemblyAI, accessed April 28, 2025, https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction
85. Introduction to Diffusion Models for Machine Learning | SuperAnnotate, accessed April 28, 2025, https://www.superannotate.com/blog/diffusion-models
86. How diffusion models work: the math from scratch | AI Summer, accessed April 28, 2025, https://theaisummer.com/diffusion-models/
87. NeurIPS Poster Diffusion for World Modeling: Visual Details Matter in Atari, accessed April 28, 2025, https://neurips.cc/virtual/2024/poster/95428
88. Your Diffusion Model is Secretly a Noise Classifier and Benefits from Contrastive Training, accessed April 28, 2025, https://neurips.cc/virtual/2024/poster/95188
89. When training a Diffusion model, what determines that the model is successful? - Reddit, accessed April 28, 2025, https://www.reddit.com/r/learnmachinelearning/comments/1fud6cn/when_training_a_diffusion_model_what_determines/
90. Making Diffusion Models Practical: A Survey on Acceleration and Optimization - Preprints.org, accessed April 28, 2025, https://www.preprints.org/frontend/manuscript/3ef2013c1dad47746fcc56736b8a27d4/download_pub
91. Deep Dive into Scalable Diffusion Models with Transformers - GitHub, accessed April 28, 2025, https://github.com/neobundy/Deep-Dive-Into-AI-With-MLX-PyTorch/blob/master/deep-dives/018-diffusion-transformer/README.md
92. Diffusion Transformer (DiT) Models: A Beginner's Guide - Encord, accessed April

28, 2025, https://encord.com/blog/diffusion-models-with-transformers/
93. [2212.09748] Scalable Diffusion Models with Transformers - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2212.09748
94. Rethinking How to Train Diffusion Models | NVIDIA Technical Blog, accessed April 28, 2025, https://developer.nvidia.com/blog/rethinking-how-to-train-diffusion-models/
95. Improving Training Efficiency of Diffusion Models via Multi-Stage Framework and Tailored Multi-Decoder Architecture, accessed April 28, 2025, https://openaccess.thecvf.com/content/CVPR2024/papers/Zhang_Improving_Training_Efficiency_of_Diffusion_Models_via_Multi-Stage_Framework_and_CVPR_2024_paper.pdf
96. An Introduction to Diffusion Models and Stable Diffusion - Marvik - Blog, accessed April 28, 2025, https://blog.marvik.ai/2023/11/28/an-introduction-to-diffusion-models-and-stable-diffusion/
97. Training Your Own Diffusion Models: A Comprehensive Guide - Algorithm Examples, accessed April 28, 2025, https://blog.algorithmexamples.com/stable-diffusion/stable-diffusion-custom-model-training-guide-2/
98. Diffusion Models: A Practical Guide - Scale AI, accessed April 28, 2025, https://scale.com/guides/diffusion-models-guide
99. Train a diffusion model - Hugging Face, accessed April 28, 2025, https://huggingface.co/docs/diffusers/en/tutorials/basic_training
100. Advanced Topics in Diffusion Models - van der Schaar Lab, accessed April 28, 2025, https://www.vanderschaar-lab.com/advanced-topics-in-diffusion-models/
101. Do diffusion models take a long time to train? - AI Stack Exchange, accessed April 28, 2025, https://ai.stackexchange.com/questions/43012/do-diffusion-models-take-a-long-time-to-train
102. Top 6 Research Papers On Diffusion Models For Image Generation - TOPBOTS, accessed April 28, 2025, https://www.topbots.com/research-papers-diffusion-models/
103. [2209.00796] Diffusion Models: A Comprehensive Survey of Methods and Applications, accessed April 28, 2025, https://arxiv.org/abs/2209.00796
104. A Survey on Generative Diffusion Models - IEEE Computer Society, accessed April 28, 2025, https://www.computer.org/csdl/journal/tk/2024/07/10419041/1UdImel27mg
105. YangLing0818/Diffusion-Models-Papers-Survey-Taxonomy - GitHub, accessed April 28, 2025, https://github.com/YangLing0818/Diffusion-Models-Papers-Survey-Taxonomy
106. [2112.10752] High-Resolution Image Synthesis with Latent Diffusion Models - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2112.10752
107. A Survey on Diffusion Models for Anomaly Detection - arXiv, accessed April 28, 2025, https://arxiv.org/html/2501.11430v3
108. accessed January 1, 1970, https://arxiv.org/abs/2501.11430v3

109. [2504.16081] Survey of Video Diffusion Models: Foundations, Implementations, and Applications - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2504.16081
110. accessed January 1, 1970, https://arxiv.org/pdf/2504.16081
111. Diffusions in Architecture: Artificial Intelligence and Image Generators - Amazon.com, accessed April 28, 2025, https://www.amazon.com/Diffusions-Architecture-Artificial-Intelligence-Generators/dp/1394191774
112. From Abstract Noise to Architectural Form: Designing Diffusion Models for Efficient Floor Plan Generation | OpenReview, accessed April 28, 2025, https://openreview.net/forum?id=skJLOae8ew
113. Generating Daylight-driven Architectural Design via Diffusion Models - Pengzhi Li, accessed April 28, 2025, https://zrealli.github.io/DDADesign/
114. New AI models possible game-changers within protein science and healthcare, accessed April 28, 2025, https://www.sciencedaily.com/releases/2025/03/250331122207.htm
115. Diffusion models for 3D generation: A survey - SciOpen, accessed April 28, 2025, https://www.sciopen.com/article/10.26599/CVM.2025.9450452
116. Nonlinear Processes in Geosciences | NP Paper of the Month: “Representation learning with unconditional denoising diffusion models for dynamical systems” - EGU Blogs, accessed April 28, 2025, https://blogs.egu.eu/divisions/np/2024/10/24/np-paper-of-the-month-representation-learning-with-unconditional-denoising-diffusion-models-for-dynamical-systems/
117. diffusion models for robotic manipulation:asurvey - arXiv, accessed April 28, 2025, https://arxiv.org/pdf/2504.08438?
118. A collection of awesome video generation studies. - GitHub, accessed April 28, 2025, https://github.com/AlonzoLeeeooo/awesome-video-generation
119. 52CV/CVPR-2024-Papers - GitHub, accessed April 28, 2025, https://github.com/52CV/CVPR-2024-Papers
120. CVPR Poster MotionEditor: Editing Video Motion via Content-Aware Diffusion, accessed April 28, 2025, https://cvpr.thecvf.com/virtual/2024/poster/29611
121. VEU-Bench: Towards Comprehensive Understanding of Video Editing - CVPR 2025, accessed April 28, 2025, https://cvpr.thecvf.com/virtual/2025/poster/34180
122. Video Editing | Papers With Code, accessed April 28, 2025, https://paperswithcode.com/task/video-editing
123. VideoHandles: Editing 3D Object Compositions in Videos Using Video Generative Priors, accessed April 28, 2025, https://openreview.net/forum?id=IReyEK7Sst
124. wentianli/awesome-video-editing: A paper list of automatic ... - GitHub, accessed April 28, 2025, https://github.com/wentianli/awesome-video-editing
125. arXiv:2504.14335v1 [cs.CV] 19 Apr 2025, accessed April 28, 2025, https://arxiv.org/pdf/2504.14335
126. Video Editing | Papers With Code, accessed April 28, 2025, https://paperswithcode.com/task/video-editing/latest
127. ECCV Conference Papers - European Computer Vision Association, accessed

April 28, 2025, https://www.ecva.net/papers.php
128. Publications - Foo Lin Geng, accessed April 28, 2025, https://lingeng.foo/publications/
129. IJCV Special Issue: Call for Papers, accessed April 28, 2025, https://avgen123.github.io/
130. Program - WACV 2024 - The Computer Vision Foundation, accessed April 28, 2025, https://wacv2024.thecvf.com/program/
131. NeurIPS 2024 Papers, accessed April 28, 2025, https://nips.cc/virtual/2024/papers.html
132. ICLR 2025 Spotlights, accessed April 28, 2025, https://iclr.cc/virtual/2025/events/spotlight-posters
133. ICLR Poster Warm Diffusion: Recipe for Blur-Noise Mixture Diffusion Models - ICLR 2025, accessed April 28, 2025, https://iclr.cc/virtual/2025/poster/28184
134. ICLR 2025 Accepted Paper List - Paper Copilot, accessed April 28, 2025, https://staging-dapeng.papercopilot.com/paper-list/iclr-paper-list/iclr-2025-paper-list/
135. Papers | MI², accessed April 28, 2025, https://www.mi2.ai/papers.html
136. NeurIPS Text-to-Audio Generation via Bridging Audio Language Model and Latent Diffusion, accessed April 28, 2025, https://neurips.cc/virtual/2024/105743
137. [2411.07765] Novel View Synthesis with Pixel-Space Diffusion Models - arXiv, accessed April 28, 2025, https://arxiv.org/abs/2411.07765
138. arXiv:2504.01521v1 [cs.LG] 2 Apr 2025, accessed April 28, 2025, https://arxiv.org/pdf/2504.01521