动作生成综述

https://github.com/haofanwang/awesome-conditional-content-generation

https://hub.baai.ac.cn/view/32620

https://blog.csdn.net/Arachis_X/article/details/136691012

论文

人体动作生成的需求通常包括一个条件信号，如**文本描述、背景音频或周围环境**，如图1所示。因此生成的动作不仅本身应该合理，还应与条件信号协调一致。

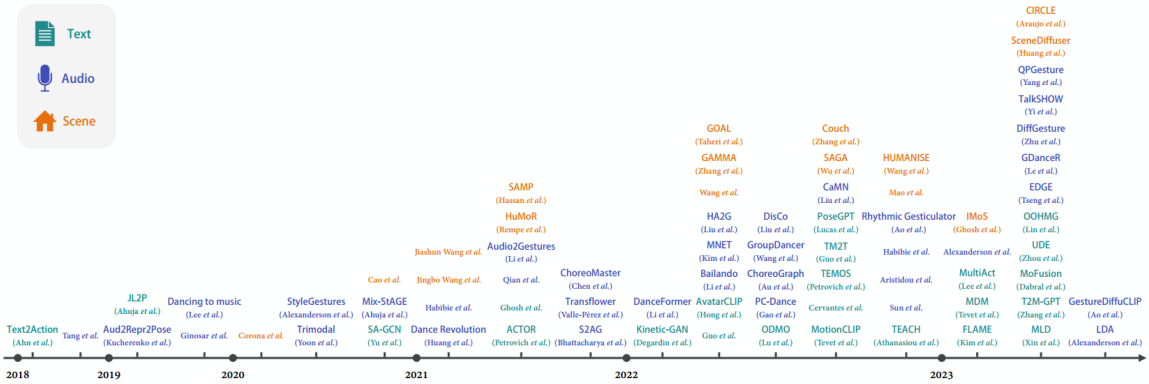
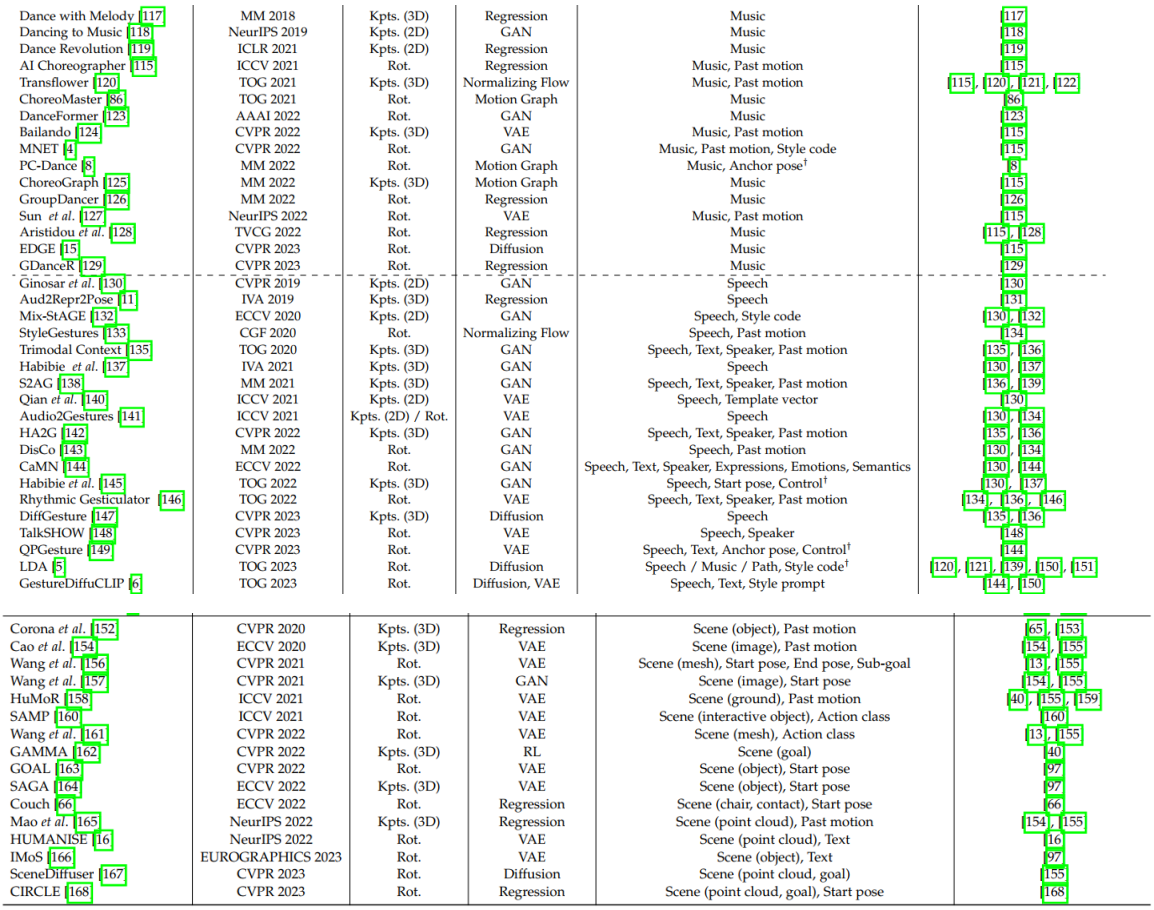


Fig. 2: Recent advances of human motion generation methods with different conditions.

Method	Venue	Representation	Model	Condition	Dataset
Action2Motion [7]	MM 2020	Rot.	VAE	Action class	[7], [87]
SA-GCN [88]	ECCV 2020	Kpts. (2D, 3D)	GAN	Action class	[87], [89]
ACTOR [90]	ICCV 2021	Rot.	VAE	Action class	[7], [87], [91]
Kinetic-GAN [92]	WACV 2022	Kpts. (3D)	GAN	Action class	[87], [89], [93]
ODMO [94]	MM 2022	Kpts. (3D)	VAE	Action class, Trajectory†	[7], [91]
PoseGPT [95]	ECCV 2022	Rot.	VAE	Action class, Duration, Past motion†	[7], [96], [97]
Cervantes *et al.* [98]	ECCV 2022	Rot.	Regression	Action class	[7], [87], [91]
MultiAct [99]	AAAI 2023	Kpts. (3D) / Rot.	VAE	Action class, Past motion†	[96]
MDM [14]	ICLR 2023	Kpts. (3D) / Rot.	Diffusion	Action class / Text	[3], [7], [91], [100]
MLD [101]	CVPR 2023	Kpts. (3D) / Rot.	Diffusion, VAE	Action class / Text	[3], [7], [91], [100]
Text2Action [102]	ICRA 2018	Kpts. (3D)	GAN	Text	[103]
JL2P [104]	3DV 2019	Kpts. (3D)	Regression	Text	[100]
Ghosh *et al.* [105]	ICCV 2021	Kpts. (3D)	GAN	Text	[100]
Guo *et al.* [3]	CVPR 2022	Kpts. (3D) / Rot.	VAE	Text, POS	[3], [100]
AvatarCLIP [106]	TOG 2022	Rot.	VAE	Text	[40]
MotionCLIP [107]	ECCV 2022	Rot.	Regression	Text	[96]
TEMOS [108]	ECCV 2022	Kpts. (3D) / Rot.	VAE	Text	[100]
TM2T [109]	ECCV 2022	Kpts. (3D) / Rot.	VAE	Text	[3], [100]
TEACH [110]	3DV 2022	Rot.	VAE	Text, Past motion†	[96]
FLAME [111]	AAAI 2023	Kpts. (3D) / Rot.	Diffusion	Text	[3], [96], [100]
T2M-GPT [112]	CVPR 2023	Kpts. (3D) / Rot.	VAE	Text	[3], [100]
OOHMG [113]	CVPR 2023	Rot.	VAE	Text	[40], [96]
UDE [114]	CVPR 2023	Rot.	Diffusion, VAE	Text / Music	[3], [115]
MoFusion [116]	CVPR 2023	Kpts. (3D)	Diffusion	Text / Music	[3], [96], [115]



Dance with Melody [117]	MM 2018	Kpts. (3D)	Regression	Music	[117]
Dancing to Music [118]	NeurIPS 2019	Kpts. (2D)	GAN	Music	[118]
Dance Revolution [119]	ICLR 2021	Kpts. (2D)	Regression	Music	[119]
AI Choreographer [115]	ICCV 2021	Rot.	Regression	Music, Past motion	[115]
Transflower [120]	TOG 2021	Kpts. (3D)	Normalizing Flow	Music, Past motion	[115], [120], [121], [122]
ChoreoMaster [86]	TOG 2021	Rot.	Motion Graph	Music	[86]
DanceFormer [123]	AAAI 2022	Rot.	GAN	Music	[123]
Bailando [124]	CVPR 2022	Kpts. (3D)	VAE	Music, Past motion	[115]
MNET [4]	CVPR 2022	Rot.	GAN	Music, Past motion, Style code	[115]
PC-Dance [8]	MM 2022	Rot.	Motion Graph	Music, Anchor pose†	[8]
ChoreoGraph [125]	MM 2022	Kpts. (3D)	Motion Graph	Music	[115]
GroupDancer [126]	MM 2022	Rot.	Regression	Music	[126]
Sun *et al.* [127]	NeurIPS 2022	Rot.	VAE	Music, Past motion	[115]
Aristidou *et al.* [128]	TVCG 2022	Rot.	Regression	Music	[115], [128]
EDGE [15]	CVPR 2023	Rot.	Diffusion	Music	[115]
GDanceR [129]	CVPR 2023	Rot.	Regression	Music	[129]
Ginosar *et al.* [130]	CVPR 2019	Kpts. (2D)	GAN	Speech	[130]
Aud2Repr2Pose [11]	IVA 2019	Kpts. (3D)	Regression	Speech	[131]
Mix-StAGE [132]	ECCV 2020	Kpts. (2D)	GAN	Speech, Style code	[130], [132]
StyleGestures [133]	CGF 2020	Rot.	Normalizing Flow	Speech, Past motion	[134]
Trimodal Context [135]	TOG 2020	Kpts. (3D)	GAN	Speech, Text, Speaker, Past motion	[135], [136]
Habibie *et al.* [137]	IVA 2021	Kpts. (3D)	GAN	Speech	[130], [137]
S2AG [138]	MM 2021	Kpts. (3D)	GAN	Speech, Text, Speaker, Past motion	[136], [139]
Qian *et al.* [140]	ICCV 2021	Kpts. (2D)	VAE	Speech, Template vector	[130]
Audio2Gestures [141]	ICCV 2021	Kpts. (2D) / Rot.	VAE	Speech	[130], [134]
HA2G [142]	CVPR 2022	Kpts. (3D)	GAN	Speech, Text, Speaker, Past motion	[135], [136]
DisCo [143]	MM 2022	Rot.	GAN	Speech, Past motion	[130], [134]
CaMN [144]	ECCV 2022	Rot.	GAN	Speech, Text, Speaker, Expressions, Emotions, Semantics	[130], [144]
Habibie *et al.* [145]	TOG 2022	Kpts. (3D)	GAN	Speech, Start pose, Control†	[130], [137]
Rhythmic Gesticulator [146]	TOG 2022	Rot.	VAE	Speech, Text, Speaker, Past motion	[134], [136], [146]
DiffGesture [147]	CVPR 2023	Kpts. (3D)	Diffusion	Speech	[135], [136]
TalkSHOW [148]	CVPR 2023	Rot.	VAE	Speech, Speaker	[148]
QPGesture [149]	CVPR 2023	Rot.	VAE	Speech, Text, Anchor pose, Control†	[144]
LDA [5]	TOG 2023	Rot.	Diffusion	Speech / Music / Path, Style code†	[120], [121], [139], [150], [151]
GestureDiffuCLIP [6]	TOG 2023	Rot.	Diffusion, VAE	Speech, Text, Style prompt	[144], [150]
Corona *et al.* [152]	CVPR 2020	Kpts. (3D)	Regression	Scene (object), Past motion	[65], [153]
Cao *et al.* [154]	ECCV 2020	Kpts. (3D)	VAE	Scene (image), Past motion	[154], [155]
Wang *et al.* [156]	CVPR 2021	Rot.	VAE	Scene (mesh), Start pose, End pose, Sub-goal	[13], [155]
Wang *et al.* [157]	CVPR 2021	Kpts. (3D)	GAN	Scene (image), Start pose	[154], [155]
HuMoR [158]	ICCV 2021	Rot.	VAE	Scene (ground), Past motion	[40], [155], [159]
SAMP [160]	ICCV 2021	Rot.	VAE	Scene (interactive object), Action class	[160]
Wang *et al.* [161]	CVPR 2022	Rot.	VAE	Scene (mesh), Action class	[13], [155]
GAMMA [162]	CVPR 2022	Kpts. (3D)	RL	Scene (goal)	[40]
GOAL [163]	CVPR 2022	Rot.	VAE	Scene (object), Start pose	[97]
SAGA [164]	ECCV 2022	Kpts. (3D)	VAE	Scene (object), Start pose	[97]
Couch [66]	ECCV 2022	Rot.	Regression	Scene (chair, contact), Start pose	[66]
Mao *et al.* [165]	NeurIPS 2022	Kpts. (3D)	Regression	Scene (point cloud), Past motion	[154], [155]
HUMANISE [16]	NeurIPS 2022	Rot.	VAE	Scene (point cloud), Text	[16]
IMoS [166]	EUROGRAPHICS 2023	Rot.	VAE	Scene (object), Text	[97]
SceneDiffuser [167]	CVPR 2023	Rot.	Diffusion	Scene (point cloud, goal)	[155]
CIRCLE [168]	CVPR 2023	Rot.	Regression	Scene (point cloud, goal), Start pose	[168]

文本驱动

文本驱动人体动作生成任务旨在根据自然语言描述生成人体动作序列。这类方法可以分为两类：动作标签驱动和自然语言驱动。动作标签驱动利用有限的预定义标签生成动作，而自然语言驱动则可以基于多样化的文本描述生成更丰富的动作。

*我们想要的是**自然语言驱动**

场景驱动

场景驱动人体动作生成任务旨在根据场景环境生成合理的人体动作。这类方法通常采用多阶段管线，首先预测目标位置或目标交互锚点，然后规划路径或轨迹，最后沿轨迹生成动作

05 未来展望

尽管这个领域取得了迅速的进展，但仍存在一些重大挑战需要未来的探索。有鉴于此，我们从不同角度展望了几个富有前景的未来研究方向，以期能激发人体动作生成研究的新突破。未来的工作可以从数据、语义、评价、可控性和互动性等方面进行深入研究，以实现更自然、真实、多样化的人体动作生成。

数据：收集高质量的人体动作数据具有挑战性，未来研究可以探讨使用异构数据源，通过弱监督学习方法或多模态基础模型整合数据优势。

语义：人体动作不仅仅是身体部位的运动，还具有丰富的非言语沟通功能。未来研究可以深入探讨从数据、方法和评价等方面捕捉人体动作与条件信号之间的高层语义关系。

评价：合适的动作评价指标至关重要，但具有挑战性。未来工作可以关注设计更符合人类感知且具有可解释性的客观评价指标。

可控性：生成内容的可控性在实际应用中非常重要。未来研究可以进一步探索可控性，以创造更用户友好的体验，例如交互式和细粒度编辑。

互动性：人体动作的互动性尚未得到充分探讨。未来研究可以关注人体动作生成在人际互动和人与环境互动背景下的应用，例如社交群体中的互动动作和动态可操作场景中的动作生成。