

Rising Stars in Data Science

November 13-14, 2023



Generative AI for Music and Audio

Hao-Wen (Herman) Dong

UC San Diego

About Me



Hi, I'm Herman.
I do **AI x Music** research.
I love music and movies!



B.S. in Electrical Engineering



Research Assistant



M.S. in Computer Science



Ph.D. in Computer Science (expected)

2013 - 2017

2017 - 2019

2019 - 2021

2019 - present

Summer 2019

Summer 2021

Summer 2022

Fall 2022

Winter 2023

Summer 2023

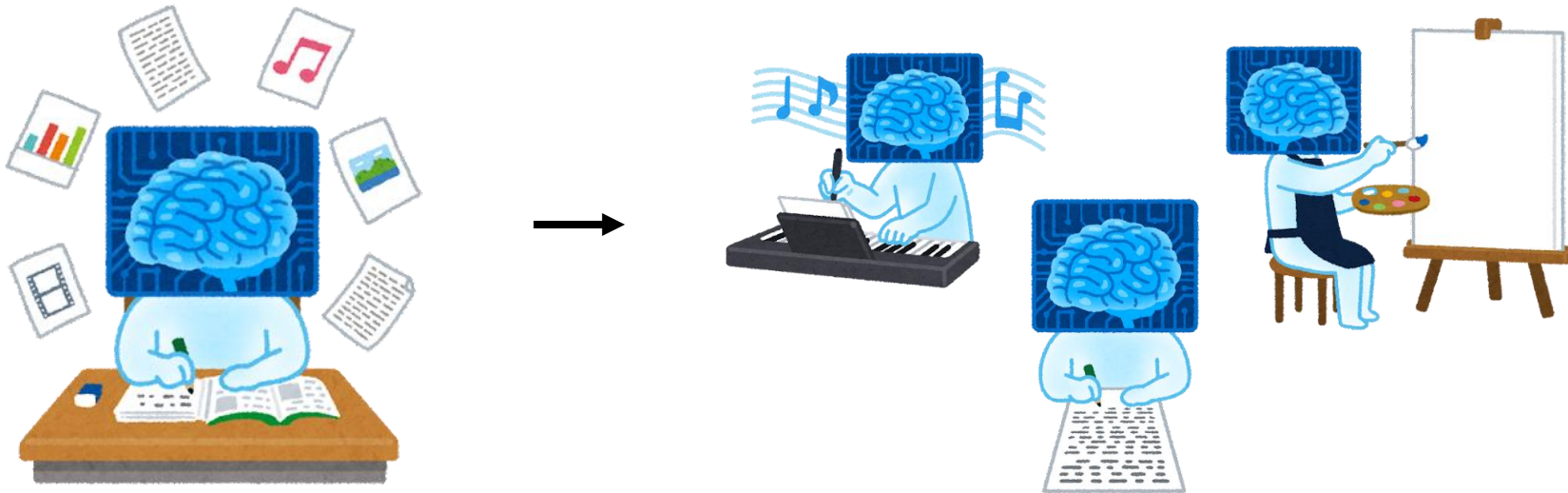
Fall 2023



Introduction

What is Generative AI?

- Generative AI is AI capable of generating text, images, or other media.



Generative AI for Visual Arts

AI made a magazine cover



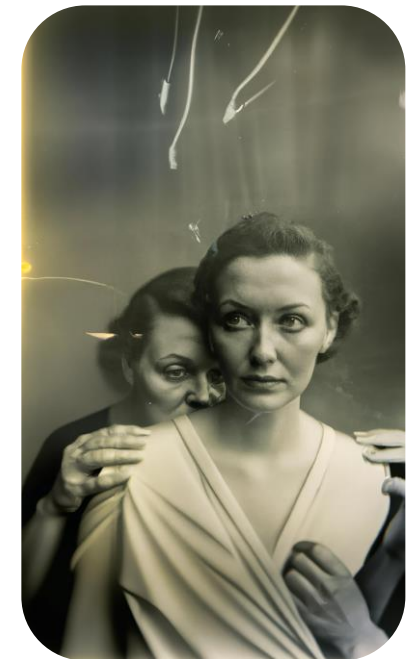
(Source: Cosmopolitan)

AI won an art contest



(Source: CNN Business)

AI won a photography contest



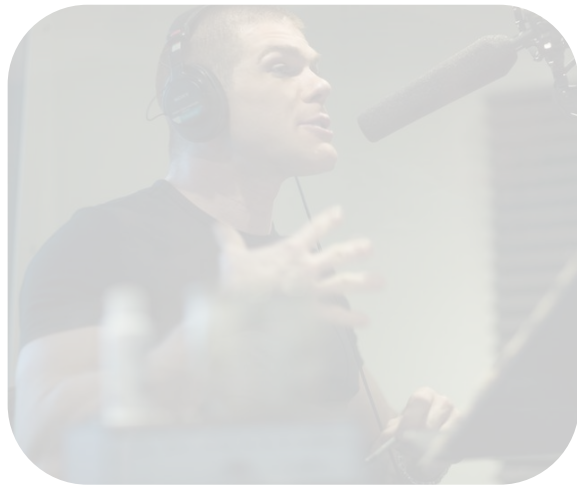
(Source: CNN)

Gloria Liu, "The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover," *Cosmopolitan*, June 21, 2022.
Rachel Metz, "AI won an art contest, and artists are furious," *CNN Business*, September 3, 2022.
Lianne Kolirin, "Artist rejects photo prize after AI-generated image wins award," *CNN*, April 18, 2023.

Types of Audio



Speech



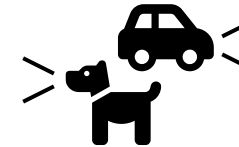
(Source: Wikimedia Commons)

Music



(Source: Wikimedia Commons)

Sound effects



(Source: Wikimedia Commons)

BPJ Media Inc, [CC BY-SA 3.0](#), via Wikimedia Commons.
Vancouver Film School Retouched version by User:Quenhitrán., [CC BY 2.0](#), via Wikimedia Commons.
The Blackbird Academy, [CC BY-SA 2.0](#), via Wikimedia Commons.
One Man Films, ["One Shot - WAR ACTION SHORT FILM," YouTube](#), September 11, 2022.

Generative AI for Music

Prompt: relaxing and smooth jazz played in a stylish cafe



Prompt: delightful country music with acoustic guitars



Prompt: cinematic and suspenseful orchestral music

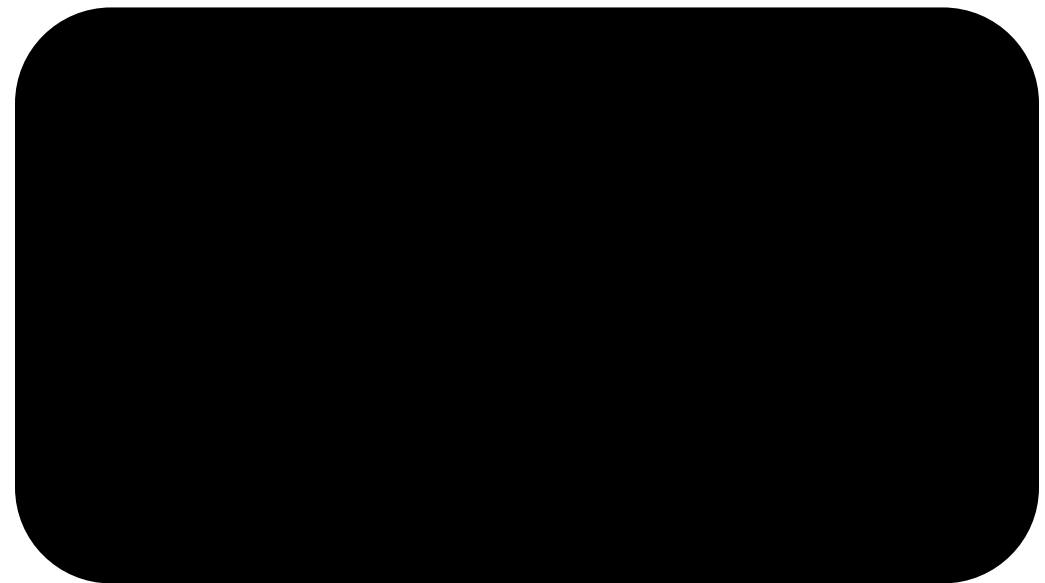


Generative AI for Sound Effects

Text-to-audio Synthesis



Image-to-audio Synthesis



My Research



Generative AI for Music & Audio



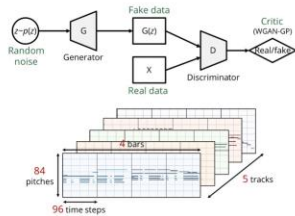
Empowering music and audio creation with machine learning

Multitrack Music Generation

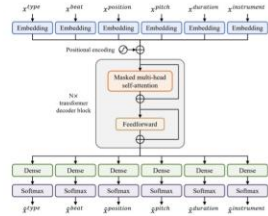
Advancing deep generative models for multitrack music



MuseGAN (AAAI 2018)



MMT (ICASSP 2023)

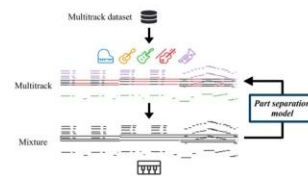


Assistive Music Creation Tools

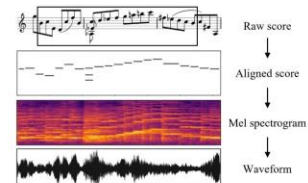
Developing AI-augmented assistive music creation tools



Arranger (ISMIR 2021)



Deep Performer (ICASSP 2022)

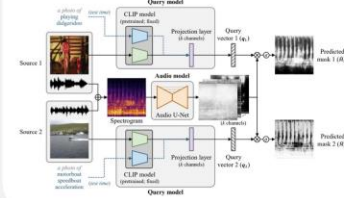


Multimodal Learning for Audio & Music

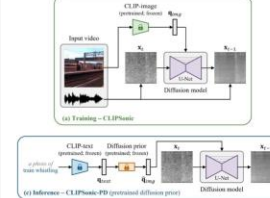
Learning sound separation and synthesis from videos



CLIPSep (ICLR 2023)



CLIPsonic (WASPAA 2023)



My Research



Generative AI for Music & Audio

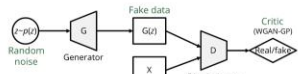
Empowering music and audio creation with machine learning

Multitrack Music Generation

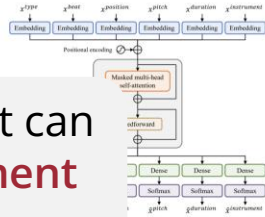
Advancing deep generative models for multitrack music



MuseGAN (AAAI 2018)



MMT (ICASSP 2023)



Assistive Music Creation Tools

Developing AI-augmented assistive music creation tools

Arranger (ISMIR 2021)



Multimodal Learning for Audio & Music



First deep neural net that can **generate multi-instrument music from scratch**

**Featured in
Amazon AWS DeepComposer**

My Research

Multitrack Music Generation

Advancing deep generative models for multitrack music



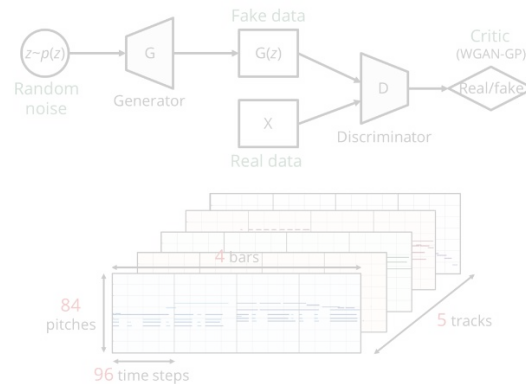
Multitrack Music Generation

Advancing deep generative models for multitrack music

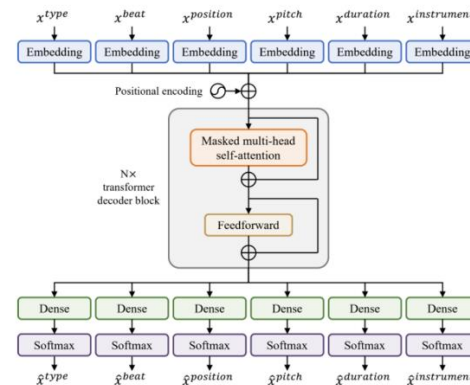
MuseGAN (AAAI 2018)



MuseGAN (AAAI 2018)



MMT (ICASSP 2023)

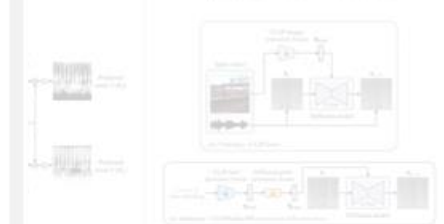


Learning for Audio & Music

separation from video



CLIPsonic (WASPAA 2023)

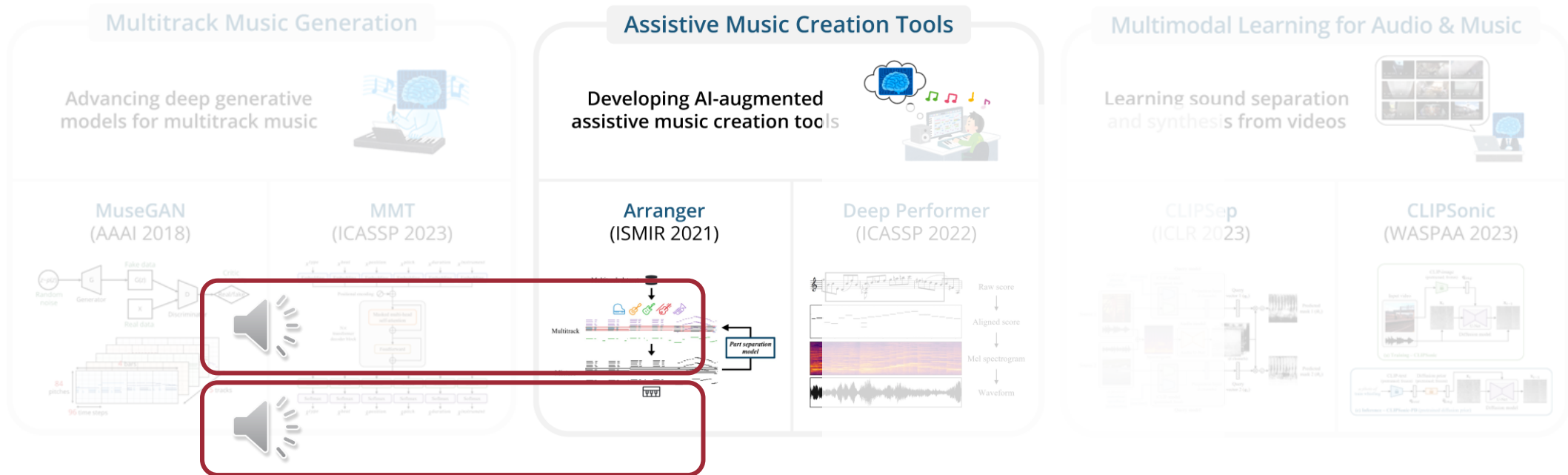


Orchestral music generation

My Research

Generative AI for Music & Audio

Empowering music and audio creation with machine learning



Automatic instrumentation

My Research

Assistive Music Creation Tools

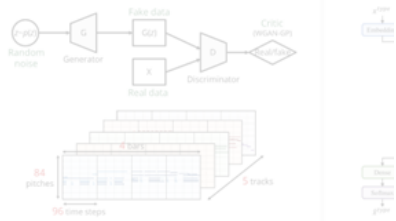
Developing AI-augmented assistive music creation tools



Multitrack Music Generation

Advancing deep generative models for multitrack music

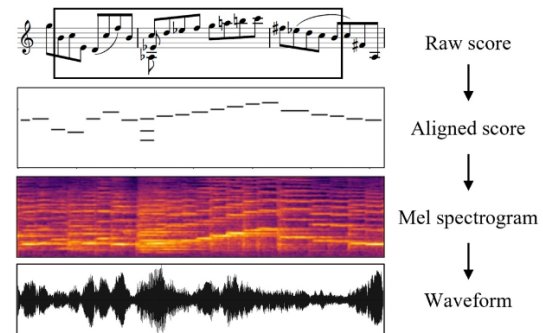
MuseGAN (AAAI 2018)



Arranger (ISMIR 2021)



Deep Performer (ICASSP 2022)



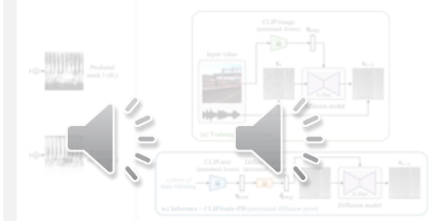
Score-to-audio synthesis

Learning for Audio & Music

Separation from videos



CLIPsonic (WASPAA 2023)



My Research



Generative AI for Music & Audio



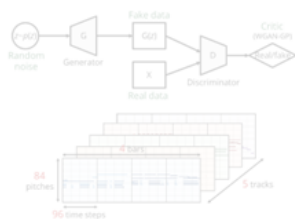
Empowering music and audio creation with machine learning

Multitrack Music Generation

Advancing deep generative models for multitrack music



MuseGAN (AAAI 2018)



MMT (ICASSP 2023)



Assistive Music Creation Tools

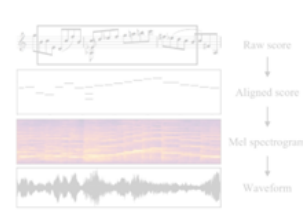
Developing AI-augmented assistive music creation tools



Arranger (ISMIR 2021)



Deep Performer (ICASSP 2022)

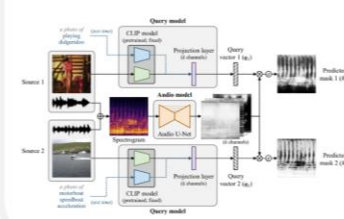


Multimodal Learning for Audio & Music

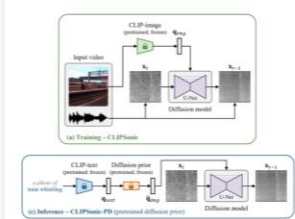
Learning sound separation and synthesis from videos



CLIPSep (ICLR 2023)



CLIPsonic (WASPAA 2023)



My Research

Multimodal Learning for Audio & Music

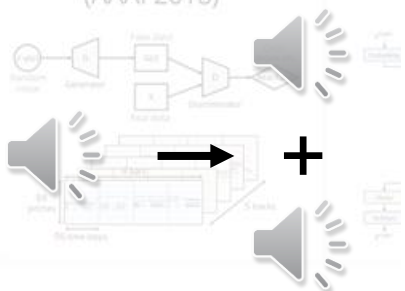
Learning sound separation and synthesis from videos



Multitrack Music Generation

Advancing deep generative models for multitrack music

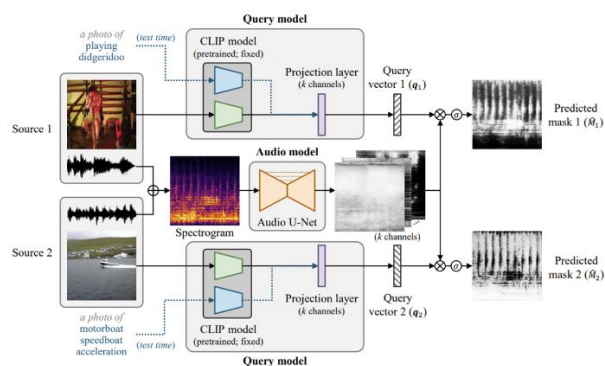
MuseGAN (AAAI 2018)



Query:

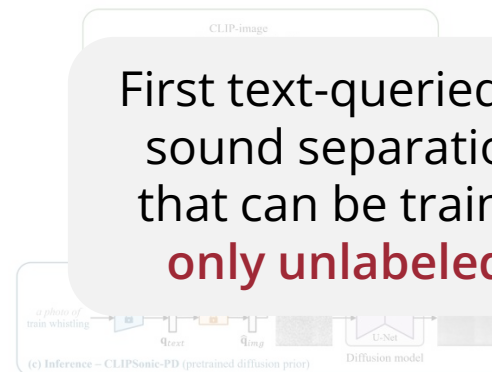
"playing harpsichord"

CLIPSep (ICLR 2023)



**Text-queried
sound separation**

CLIPSonic (WASPAA 2023)

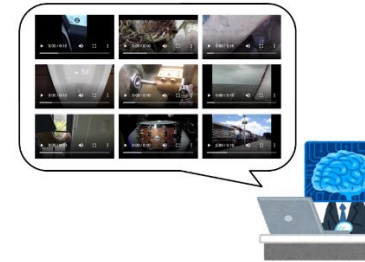


First text-queried universal sound separation model that can be trained **using only unlabeled videos**

My Research

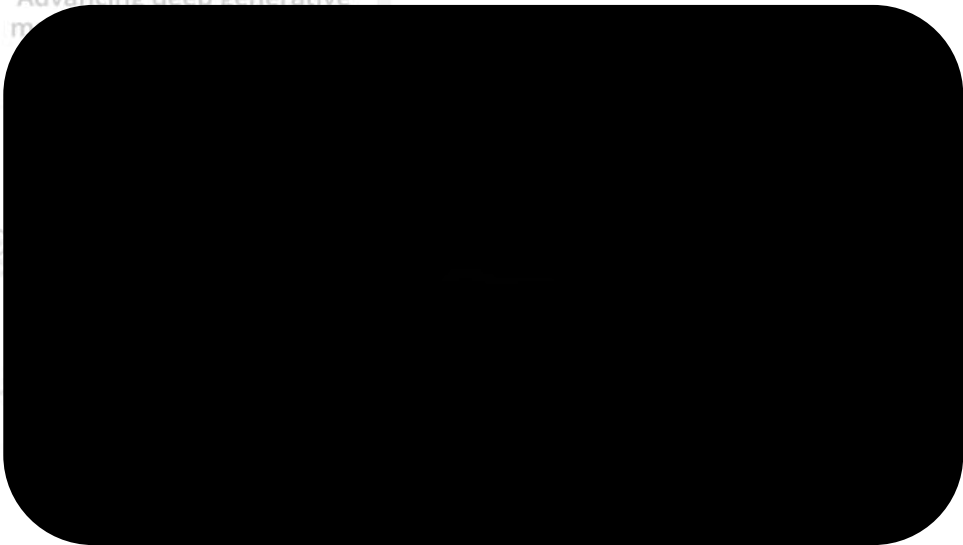
Multimodal Learning for Audio & Music

Learning sound separation and synthesis from videos



Multitrack Music Ge

Advancing deep generative

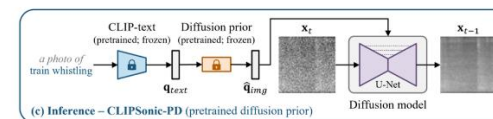
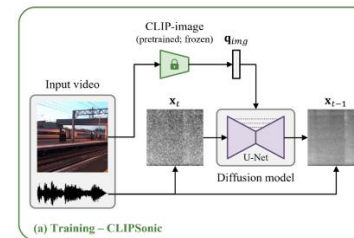


Learning for Audio & Music

Separation from videos



CLIPsonic (WASPAA 2023)

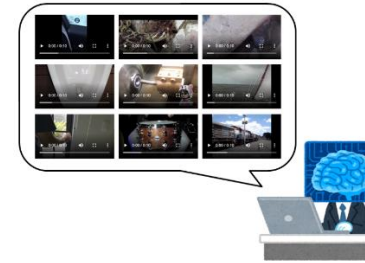


Text-to-audio synthesis

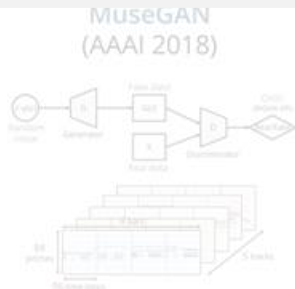
My Research

Multimodal Learning for Audio & Music

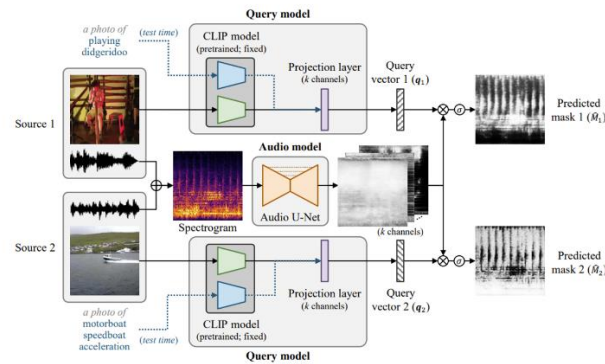
Learning sound separation and synthesis from videos



A new direction of learning audio-text correspondence by **leveraging the visual modality as a bridge**

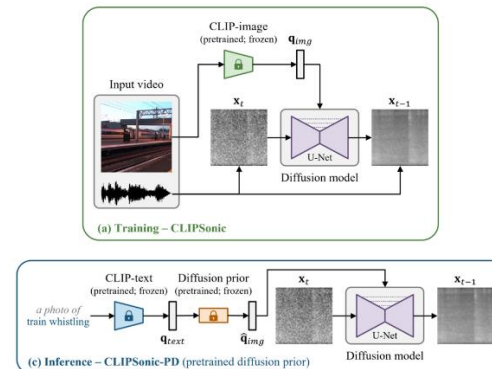


CLIPSep (ICLR 2023)



Text-queried sound separation

CLIPSonic (WASPAA 2023)



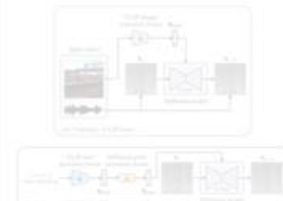
Text-to-audio synthesis

Learning for Audio & Music

Separation from videos



CLIPSonic (WASPAA 2023)



My Research



Generative AI for Music & Audio

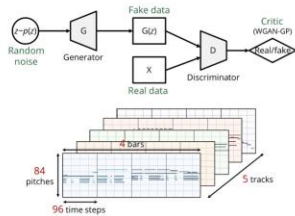
Empowering music and audio creation with machine learning

Multitrack Music Generation

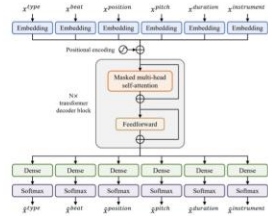
Advancing deep generative models for multitrack music



MuseGAN (AAAI 2018)



MMT (ICASSP 2023)

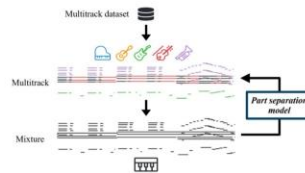


Assistive Music Creation Tools

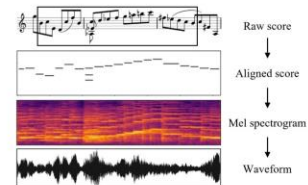
Developing AI-augmented assistive music creation tools



Arranger (ISMIR 2021)



Deep Performer (ICASSP 2022)

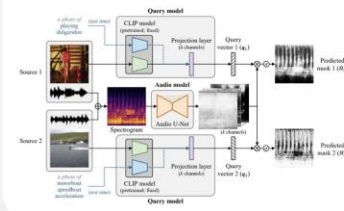


Multimodal Learning for Audio & Music

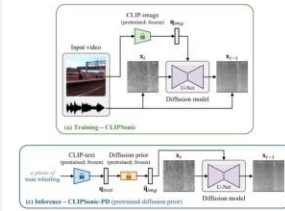
Learning sound separation and synthesis from videos



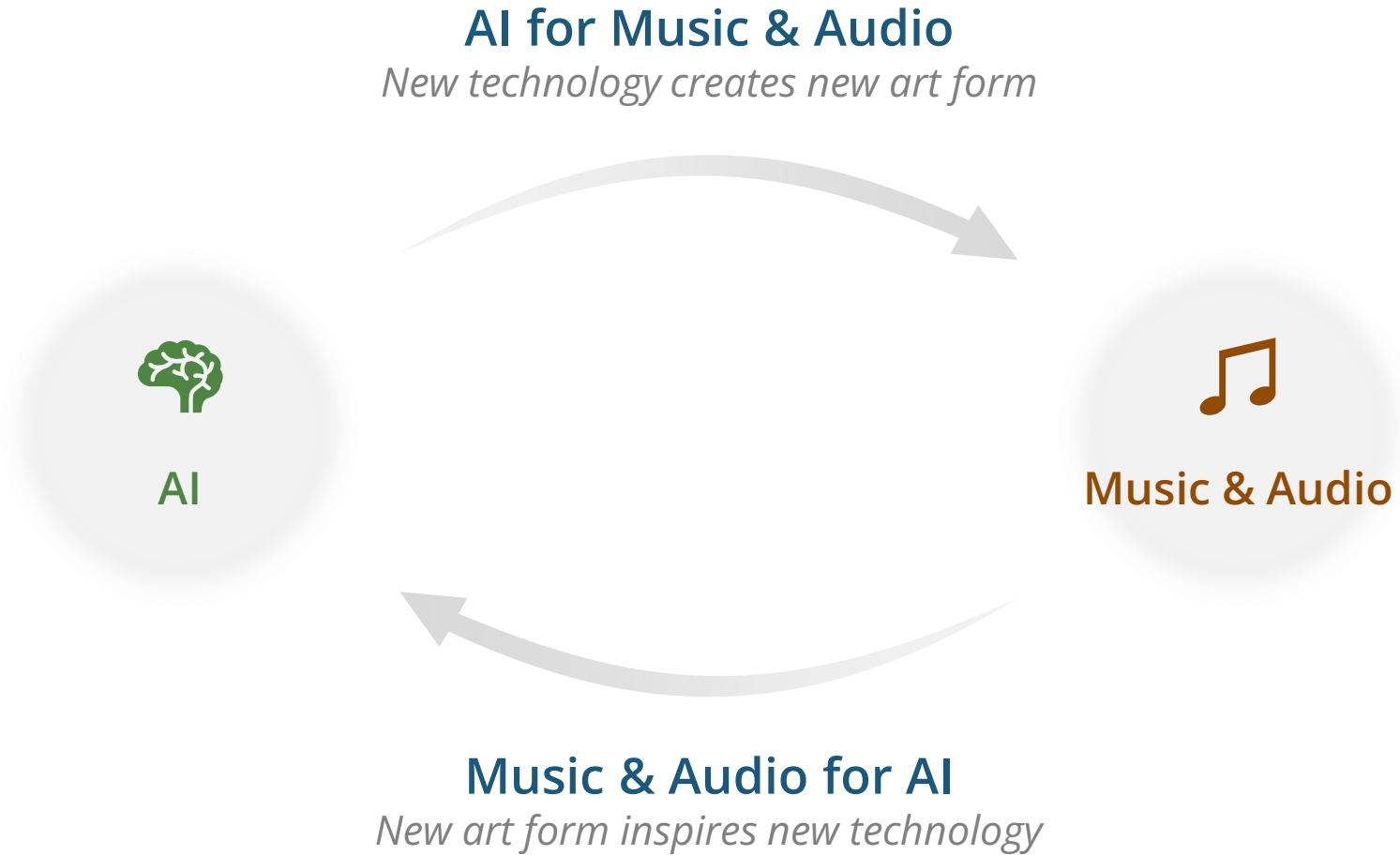
CLIPSep (ICLR 2023)



CLIPsonic (WASPAA 2023)



My Research

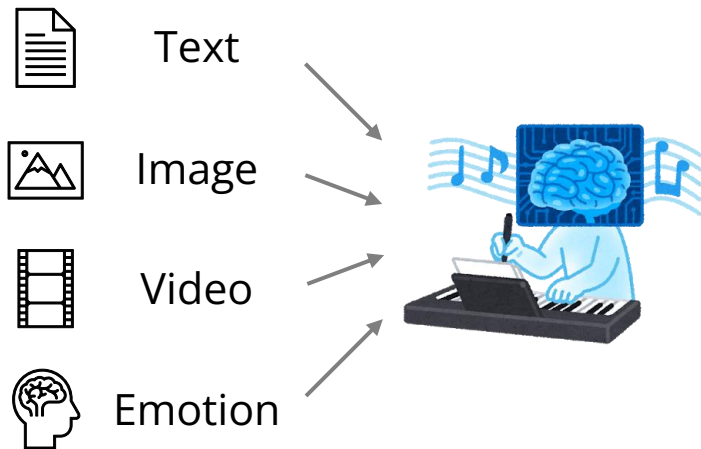


Future Directions

Future Directions

Multimodal

Multimodal generative AI with music and audio



Interactive

Interactive AI tools for music and audio production



Human-inspired

Human-like machine learning algorithms for music



Acknowledgements



Julian McAuley



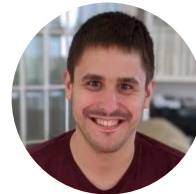
Taylor Berg-Kirkpatrick



Yi-Hsuan Yang



Shlomo Dubnov



UC San Diego



Dolby

SONY

amazon



Thank you!



Generative AI for Music & Audio



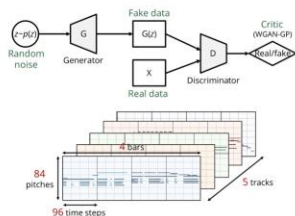
Empowering music and audio creation with machine learning

Multitrack Music Generation

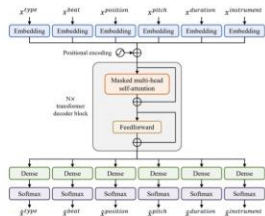
Advancing deep generative models for multitrack music



MuseGAN (AAAI 2018)



MMT (ICASSP 2023)

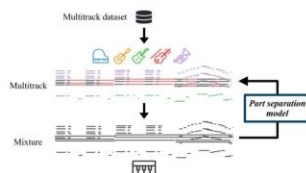


Assistive Music Creation Tools

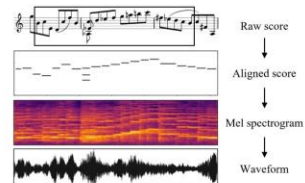
Developing AI-augmented assistive music creation tools



Arranger (ISMIR 2021)



Deep Performer (ICASSP 2022)

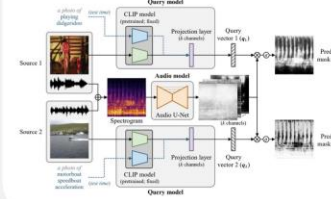


Multimodal Learning for Audio & Music

Learning sound separation and synthesis from videos



CLIPSep (ICLR 2023)



CLIPsonic (WASPAA 2023)

