Polyffusion: A Diffusion Model for Polyphonic Score Generation with Internal and External Controls

Lejun Min[124], **Junyan Jiang**[12], **Gus Xia**[12], **Jingwei Zhao**[3]

{lejun.min, junyan.jiang, gus.xia}@mbzuai.ac.ae, jzhao@u.nus.edu

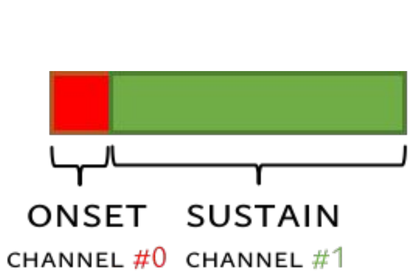
[1] MBZUAI, [2] NYU Shanghai, [3] National University of Singapore, [4] Shanghai Jiao Tong University

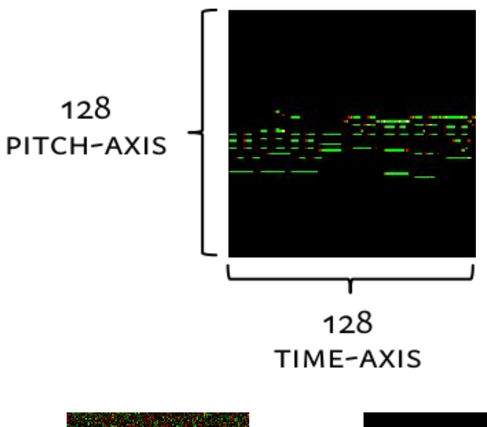
Introduction

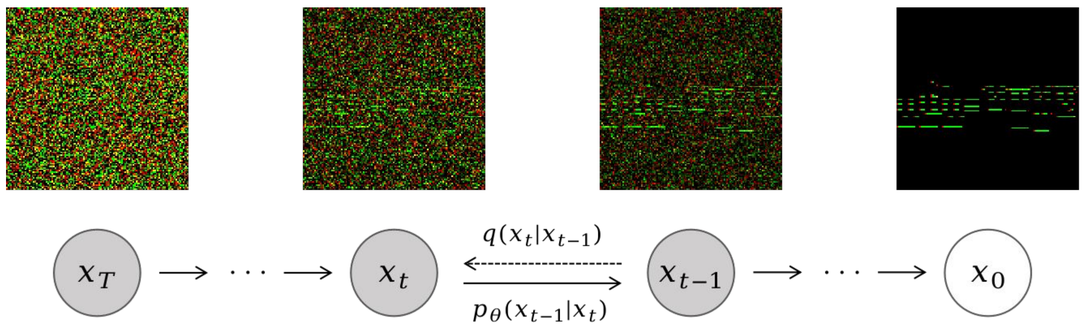
- We generate polyphonic music with diffusion models by treating piano-rolls as images.
- We unify a wide range of music creation tasks with just one model and two types of controls. The tasks include *melody generation given accompaniment*, *accompaniment generation given melody*, *music segment inpainting*, and *music arrangement given chords or textures*.

Model & Data Processing

- **Data Processing:** Transform 8-bar piano-rolls into 2-channel images (minimum time unit is 1/4 beat).
- **Training & Inference:** Implement a DDPM to reconstruct piano-roll images from Gaussian noise.



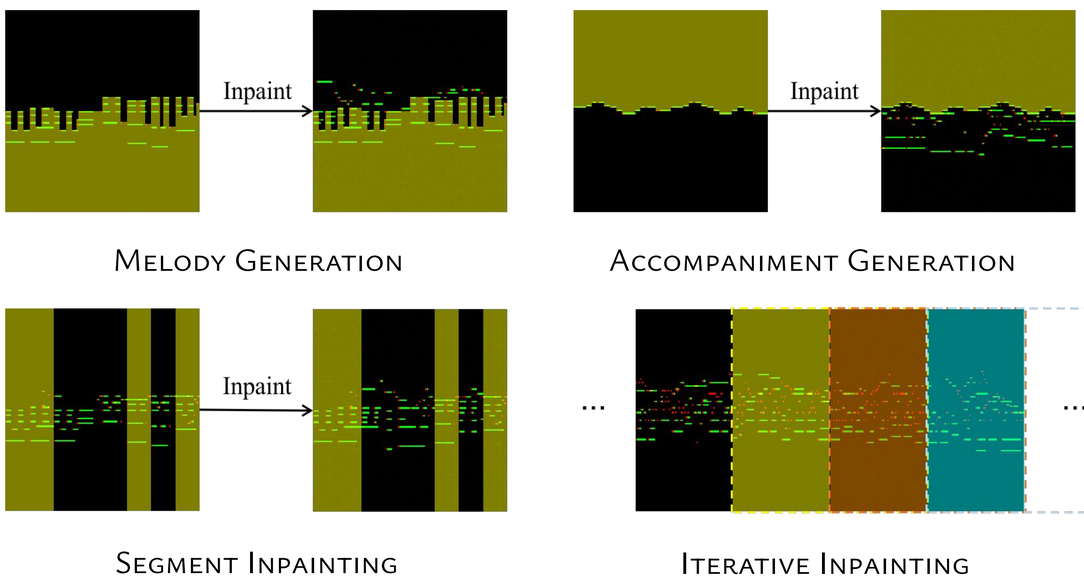




Controllability

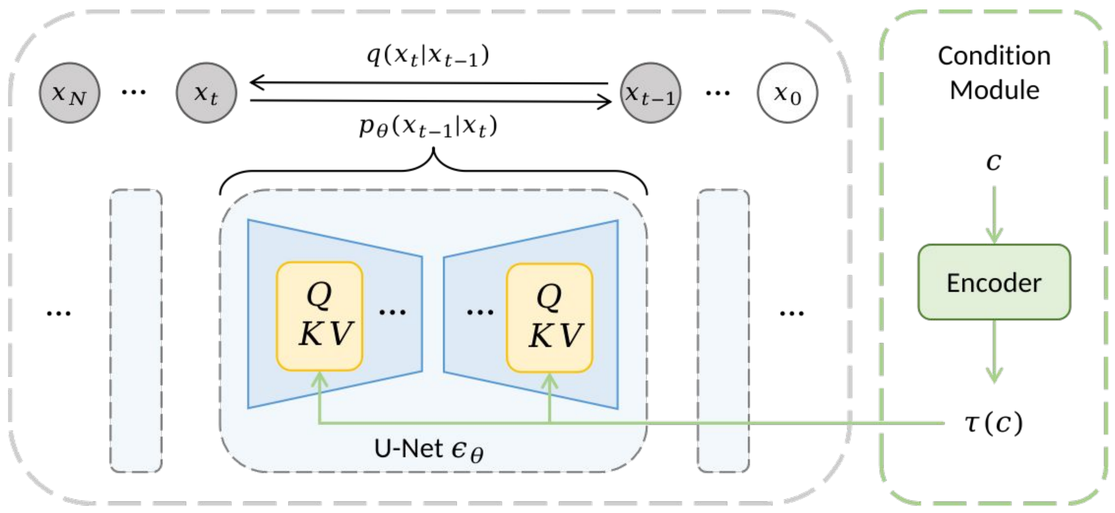
Internal Control: controls internally by pre-defined parts

- Mask out the pre-defined part and only denoise the rest.
- No separated training is needed since the control is applied post-hoc.



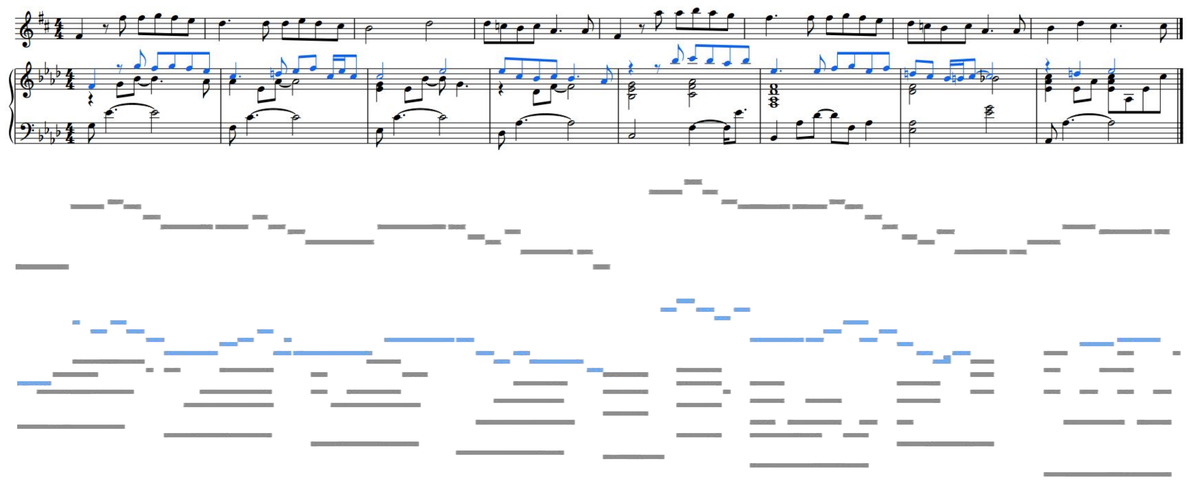
External Control: controls by external signals

- A conditioning module is added. The external condition is encoded and mapped to the cross-attention layers in the model.
- The Classifier-free Guidance (CFG) technique is applied to control the variance of the generation.



A Hybrid Control Case

Text-specified Melody Generation: The model applies internal control via inpainting for the pre-defined accompaniment, and also incorporates the texture input as an external control.

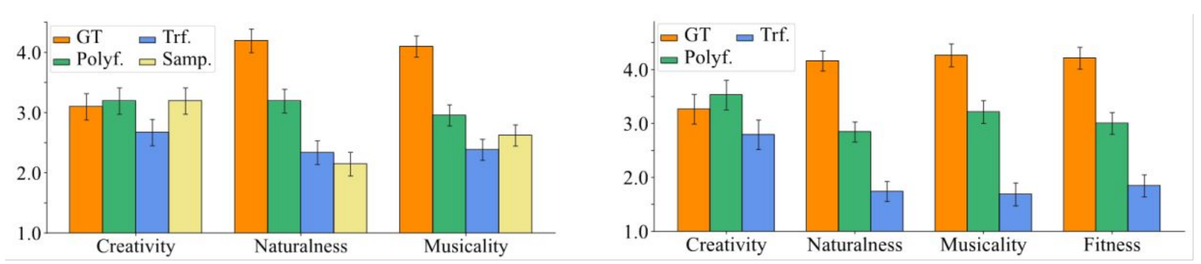


Experiments

- **Dataset:** POP909 dataset, quantized to 16[th]-note level, 8-bar segmentation with 1-bar hopping size.
- **Specifications:** 1K diffusion steps, Adam Optimizer, learning rate 5e-5.

	(1) Uncond. Gen.	(2) Acc. Gen.	(3) Seg. Inp.	(4) Chord Cond.	(5) Texture Cond.
Objective Metrics	$\mathcal{D}_\mathrm{P}$, $\mathcal{D}_\mathrm{D}$	$\mathcal{D}_\mathrm{P}$, $\mathcal{D}_\mathrm{D}$	$\mathcal{D}_\mathrm{P}$, $\mathcal{D}_\mathrm{D}$	$\mathcal{D}_\mathrm{P}$, $\mathcal{D}_\mathrm{D}$, CD	$\mathcal{D}_\mathrm{P}$, $\mathcal{D}_\mathrm{D}$, OD
Subjective Metrics	C, N, M	C, N, M, F	N/A	N/A	N/A
Generative Length	8 bars	8 bars	4 bars	8 bars	8 bars
Transformer Baselines	Wang	Wang	Chang	Wang	Wang
Sampling Baselines	Wang	N/A	Wang	Wang	Wang

	Uncond. Gen.		Acc. Gen.		Seg. Inp.		Chord Cond.			Texture Cond.		
	$\mathcal{D}_\mathrm{P}\uparrow$	$\mathcal{D}_\mathrm{D}\uparrow$	$\mathcal{D}_\mathrm{P}\uparrow$	$\mathcal{D}_\mathrm{D}\uparrow$	$\mathcal{D}_\mathrm{P}\uparrow$	$\mathcal{D}_\mathrm{D}\uparrow$	$\mathcal{D}_\mathrm{P}\uparrow$	$\mathcal{D}_\mathrm{D}\uparrow$	CD$\downarrow$	$\mathcal{D}_\mathrm{P}\uparrow$	$\mathcal{D}_\mathrm{D}\uparrow$	OD$\downarrow$
Polyffusion	**0.89**	**0.93**	**0.89**	**0.96**	**0.90**	**0.93**	0.90	**0.96**	0.75	0.88	**0.98**	1.85
Polyffusion-S5	0.89	0.93	0.89	0.96	0.90	0.93	**0.92**	0.81	**0.51**	0.87	0.97	1.75
Polyffusion-A	0.89	0.93	0.89	0.96	0.90	0.93	0.90	0.94	0.79	**0.95**	**0.98**	4.37
Transformer	0.78	0.84	0.88	0.89	**0.90**	0.83	0.87	0.88	0.56	0.84	0.93	**0.13**
Sampling	0.86	0.90	N/A	N/A	0.89	0.91	0.86	0.90	0.70	0.91	0.93	0.20



Future Work

- Performance features: velocity, and nuanced timing.
- Multi-track extension.

Demo Page & Code Repo

https://polyffusion.github.io

https://github.com/aik2mlj/polyffusion