PANDORA: Diffusion Policy Learning for Dexterous Robotic Piano Playing

Yanjia Huang, Renjie Li, Zhengzhong Tu*
Department of Computer Science and Engineering, Texas A&M University
*Corresponding author: tzz@tamu.edu

TACO-Group @ Texas A&M University

NEURAL INFORMATION PROCESSING SYSTEMS

1. Motivation & Challenge

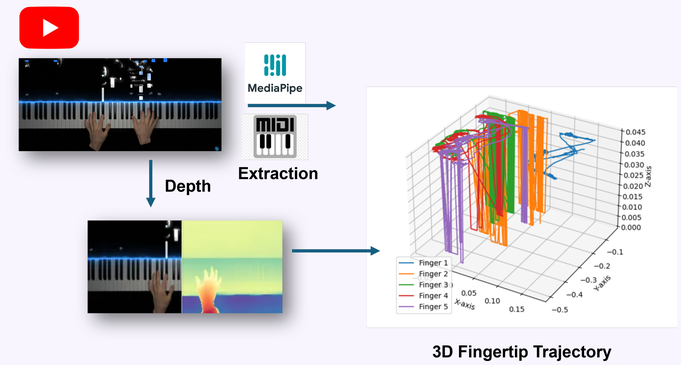
Dexterous robotic piano playing demands both **high precision** and **expressive artistry**.

Key Challenges:

- **High-dimensional Control:** 20+ DoF per hand
- **Precision vs. Expressiveness:** Balancing accuracy with musical nuances
- **Hand Coordination:** Left (rhythm) vs. Right (melody)
- **Reward Design:** Traditional metrics miss artistic qualities

Solution: Diffusion-based policy with LLM semantic feedback for musical artistry.

2. Data Preparation Pipeline



Data Preparation

Enhanced 3D Trajectory Extraction:

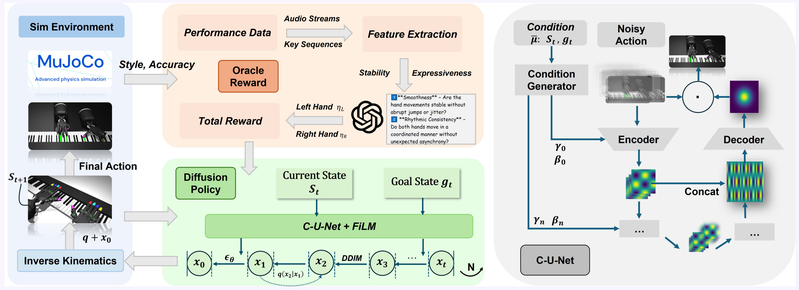
- Scrape YouTube piano performances with paired MIDI files
- Extract 2D fingertips using MediaPipe
- Augment with **DepthAnything** for depth cues
- Fuse depth + 2D to obtain robust 3D fingertip trajectories

3. Key Contributions

Novel Technical Innovations:

- **Diffusion-Based Policy:** Conditional U-Net with FiLM for smooth action generation via DDIM
- **Composite Reward:** Task accuracy + audio fidelity + style mimicry + LLM semantic evaluation
- **Hand-Specific Modulation:** Left (stability) vs. Right (expressiveness) dynamic rewards
- **Residual IK:** Combines IK solver with learned residuals for precise control
- **Enhanced Data:** DepthAnything for robust 3D fingertip trajectory extraction

4. Framework Overview



PANDORA uses a FiLM-conditioned U-Net to iteratively denoise noisy action sequences into smooth trajectories, combines them with a residual IK head, and optimizes a composite reward that includes task accuracy, audio fidelity, style mimicry, and LLM-based semantic feedback.

5. Diffusion-Based Policy Learning

DDIM Denoising: Starting from $x_T \sim \mathcal{N}(0, I)$, iteratively refine:

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1-\bar{\alpha}_{t-1}}\epsilon_\theta(x_t, t)$$

Architecture:

- U-Net with 4 blocks (64, 128, 256, 512 channels)
- FiLM layers inject state s and goal g_t
- $T = 100$ steps with cosine schedule

Training: Cosine LR ($1e{-}4$ to $1e{-}6$), EMA (0.9999), $\approx$1 hr on RTX 4090

6. Composite Reward Function

Multi-Component Design:

(1) Task $R_{\text{task}} = w_{\text{press}}(1 - \text{error}) - w_{\text{fp}} \cdot \text{FP}$

(2) Audio $R_{\text{audio}} = \dfrac{X_{\text{target}} \cdot X_{\text{robot}}}{\|X_{\text{target}}\| \, \|X_{\text{robot}}\|}$

(3) Style $R_{\text{style}} = -\|\tau_{\text{robot}} - \tau_{\text{human}}\|_2^2$

(4) LLM Oracle with Hand-Specific Modulation:

$$R^L_{\text{LLM}} = S_{\text{LLM}} \times \eta_L, \qquad R^R_{\text{LLM}} = S_{\text{LLM}} \times \eta_R$$

where η_L emphasizes stability, η_R emphasizes expressiveness.

Final Reward:

$$R = \alpha R_{\text{task}} + \beta R_{\text{audio}} + \gamma R_{\text{style}} + \delta\left(R^L_{\text{LLM}} + R^R_{\text{LLM}}\right)$$

7. Quantitative Results

13-song evaluation using Precision, Recall, and F1 metrics.

Method	Precision	Recall	F1
PianoMime (Two-stage)	0.68	0.54	0.57
PianoMime (Residual)	0.70	0.56	0.58
PANDORA	**0.78**	**0.60**	**0.68**

+10% F1 improvement with $\sim$**3**$\times$ **faster** training vs. baselines.

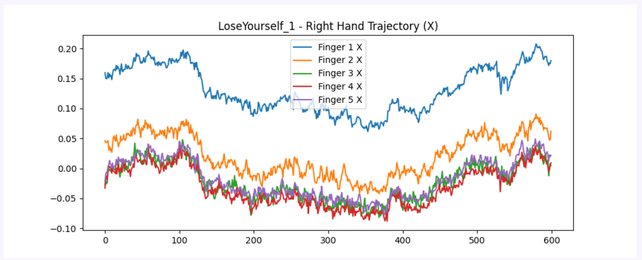
8. Ablation Study

Configuration	Mean F1 Score
LLM + Residual (Full)	**0.90**
LLM - no Residual	0.73
no LLM + Residual	0.68
no LLM - no Residual	0.62

Insights:

- **Residual policy** is essential for precise, stable key strikes.
- **LLM feedback** is critical for expressive phrasing and dynamics.
- Their **synergy** yields the best accuracy and artistry.

9. Trajectory Analysis



Conclusion & Resources

Summary: PANDORA bridges precision and artistry via diffusion, residual IK, and LLM-guided rewards.

Future: Faster sampling, multi-instrument extension, real-world deployment.

Project: https://taco-group.github.io/PANDORA
Paper: https://arxiv.org/abs/2503.14545
Contact: yanjia0812@tamu.edu

Project Page