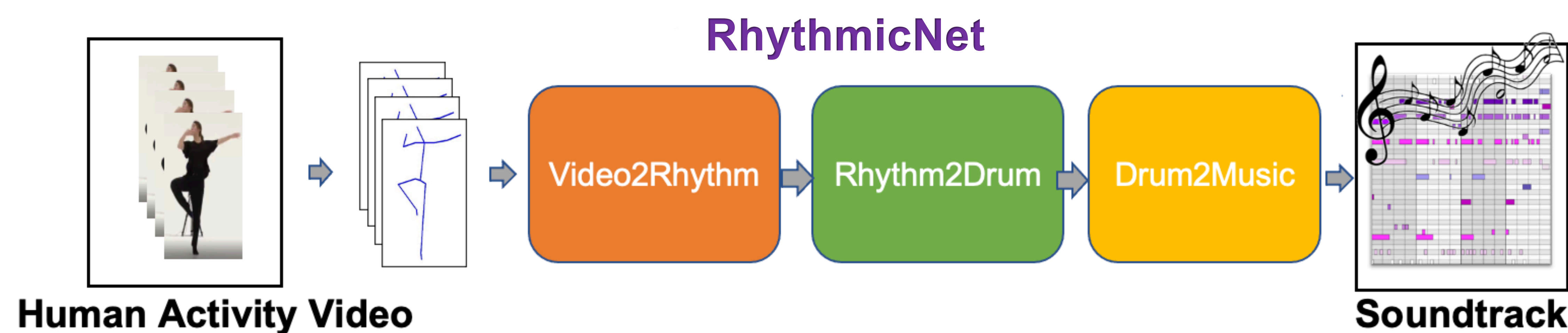


# How does it sound? Generation of Rhythmic Soundtrack for Human Movements

01

Kun Su<sup>1\*</sup>, Xiulong Liu<sup>1\*</sup>, Eli Shlizerman<sup>1,2</sup> \*Equal Contribution

<sup>1</sup> Electrical & Computer Engineering, <sup>2</sup> Applied Mathematics, University of Washington, Seattle, USA



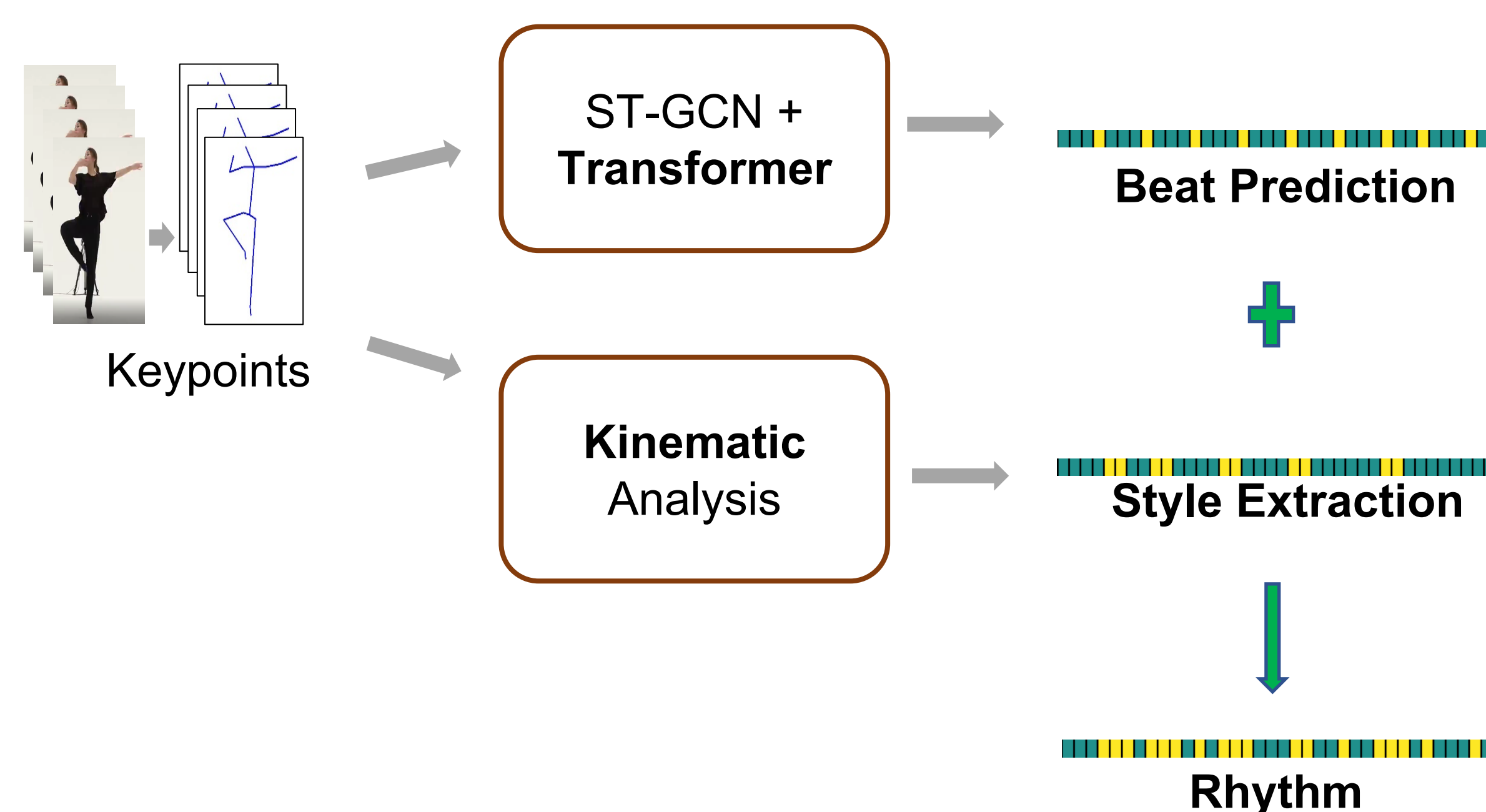
**RhythmicNet** pipeline takes as an input video with human movements and generates a soundtrack for it.

## RhythmicNet Components:

1. **Video2Rhythm**: Inferring rhythm from human activity
2. **Rhythm2Drum**: Translating rhythm to drum music
3. **Drum2Music**: Adding piano/guitar track

## Video2Rhythm

We decompose the rhythm into two streams: **Beat** and **Style**.

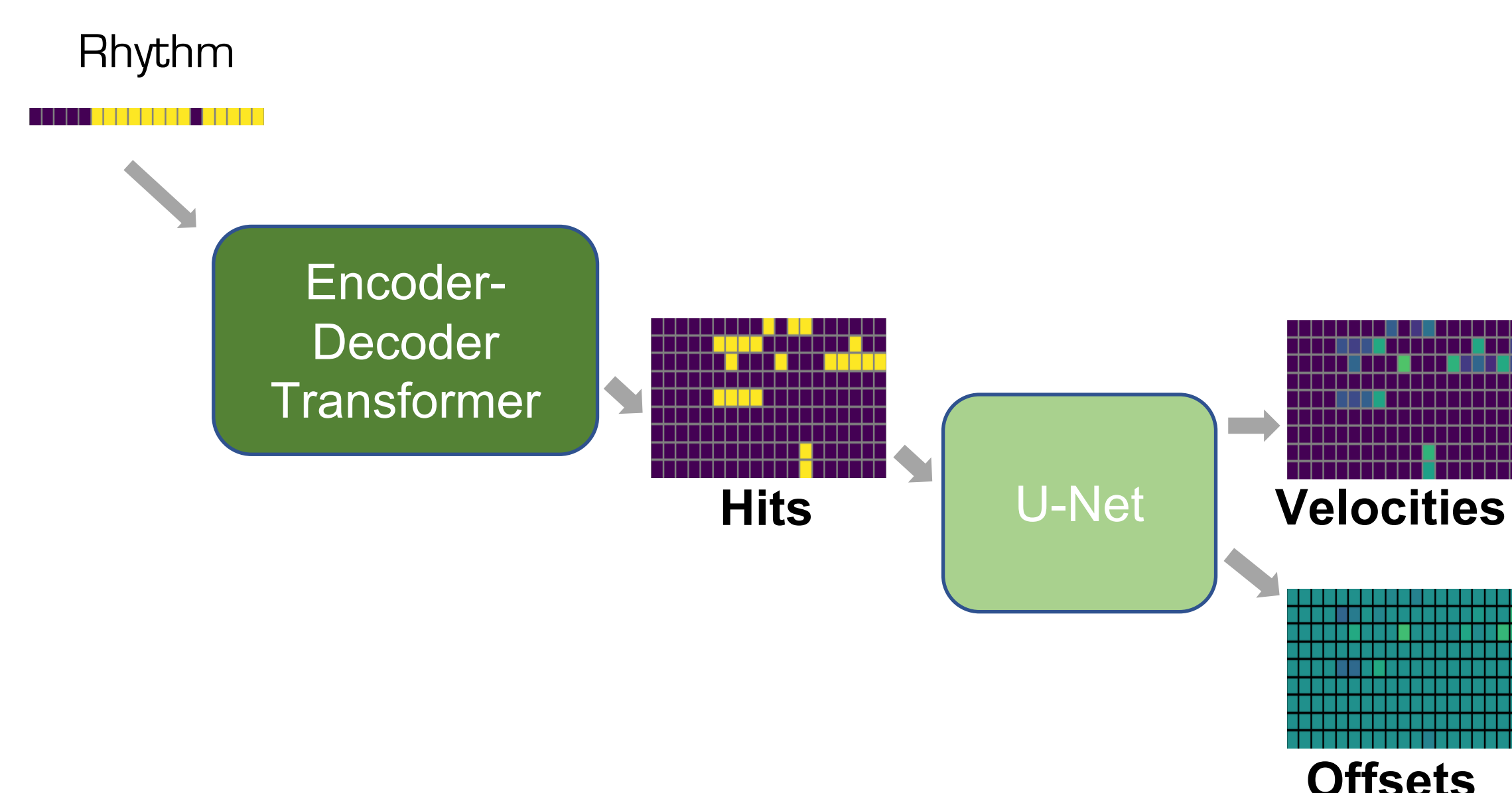


**Beat**: Body keypoints from the video are transformed to a beat signal with a novel **SpatioTemporal Graph Convolutional network** in conjunction with **Transformer Encoder**. These network models capture motion features and translate them to musical beat estimation.

**Style**: To capture various rapid movements and motion characteristics that are not reflected by the beat, we propose an additional stream, called the style. The combination of the two streams constitutes the rhythm of the dynamics.

## Rhythm2Drum

Rhythm2Drum transforms the rhythm into drum sounds. It implements an **Encoder-Decoder Transformer** that given the rhythm generates the drums performance **Hits (onsets)**. The Transformer generates natural and diverse drum onsets by autoregressively learning the hits as a word sequence conditioned on the rhythm.

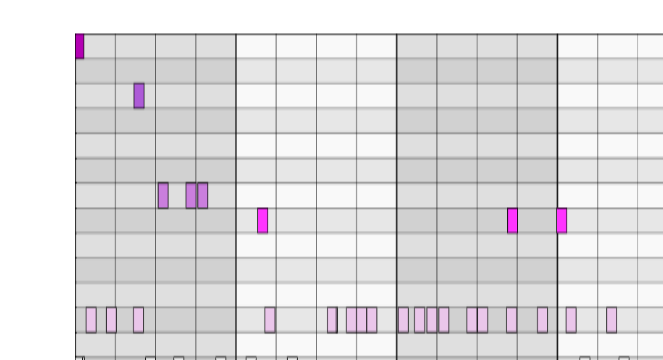


Subsequently, a **U-net** type model generates the drum **Velocities** and **Offsets**. The U-Net learns them conditioned on the **Hits** matrix.

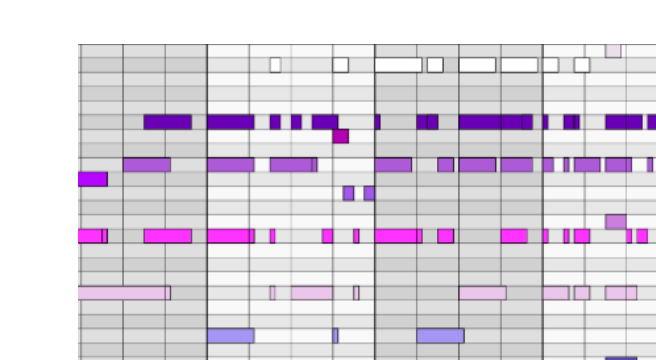
Visit our poster to see and listen to examples! Also at: <https://github.com/shlizee/RhythmicNet>

## Drum2Music

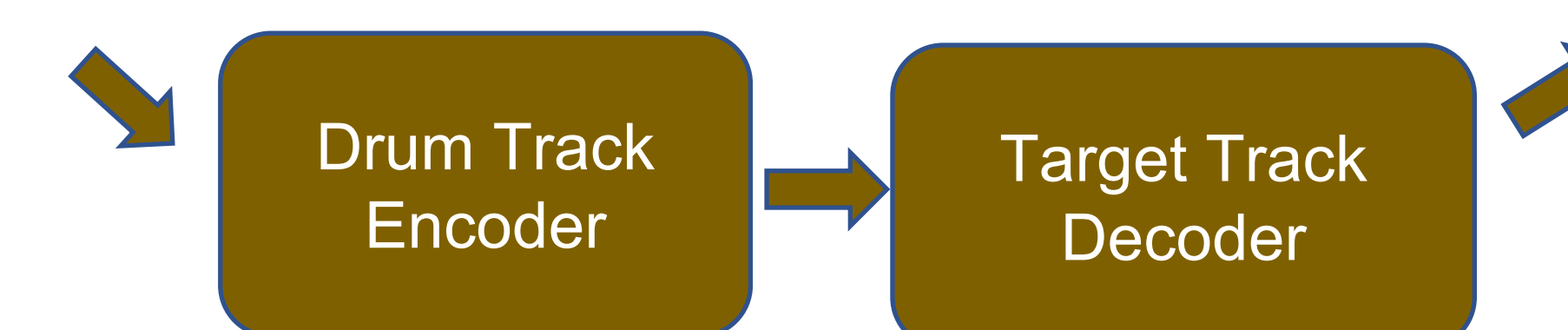
Drum2Music completes the drum music by adopting an **Encoder-Decoder** architecture using **transformer-XL** to generate a music track of either piano or guitar conditioning on the generated drum performance.



Drum Midi



Multi Instrumental Midi



## Evaluation

We evaluate each component of RhythmicNet quantitatively using available dataset (see paper for more details). We also perform a qualitative human perceptual evaluation of soundtracks of 'in the wild' videos.

### Preference

	No Soundtrack	Drums Only	Drums + Piano	Drums + Guitar
Votes	7.3%	31.2%	<b>32.1%</b>	29.4%

Soundtrack match to video			Soundtrack match to video (Ablation)		
Random	Shuffle	RhythmicNet	Random +Groove VAE	Video2Rhythm +GrooVAE	Video2Rhythm +Rhythm2Drum
30.8%	27.8%	<b>41.4%</b>	23.3%	33.3%	<b>43.4%</b>

