Lyrics & Tune Style Transfer

Qixuan Xiao / Ming Wang / Yanhao Shen / Honghu Luo / Haozhe Liu
AI BERT @ University of Southern California

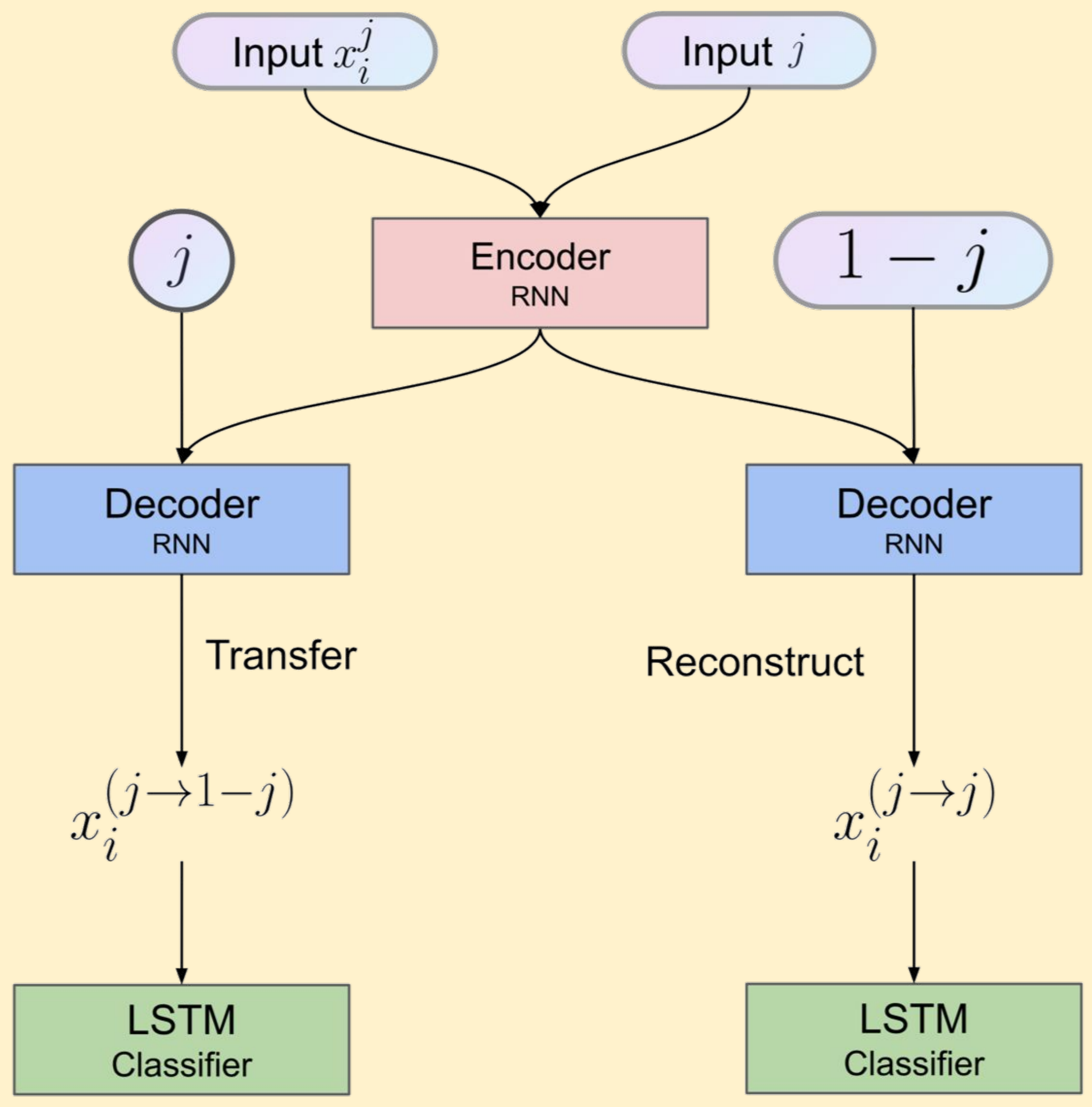
USC Viterbi
School of Engineering
Department of Computer Science

Abstract

Song lyrics are a source of rich structural and linguistic information in Music but have not been explored extensively in NLP. One challenge is that there is no set of parallel corpora to translate. In this project, we adopt the Encoder-Decoder approach to unsupervised text style transfer to solve this problem. And evaluate this model from different aspects. GANs' great success in the vision domain inspires us to adopt style transfer in the audio domain. Music is hard to be understood by machines. We apply end-to-end supervised learning with track-wise tune transfer.

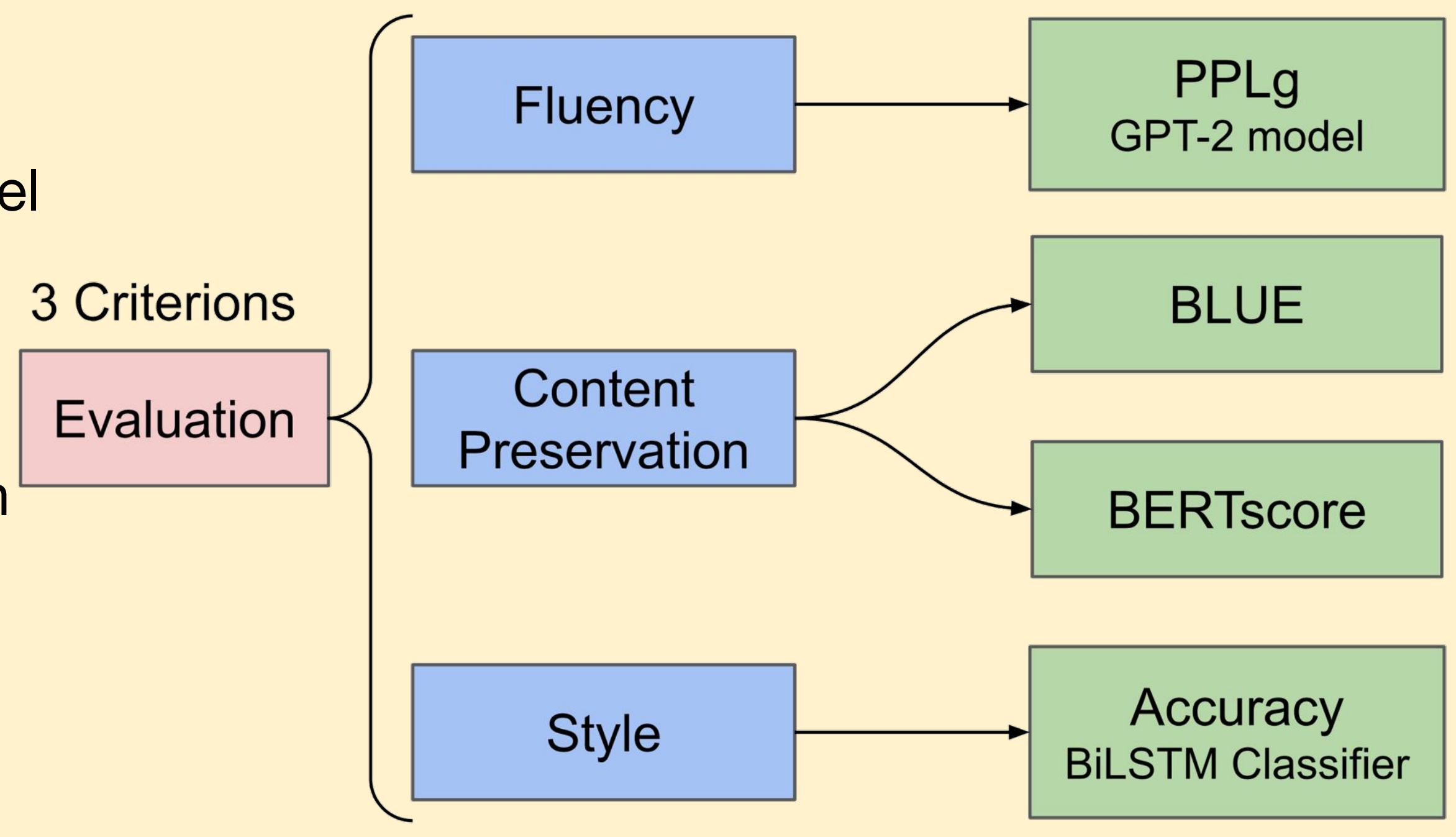
Lyrics Style Transfer

- The content of the lyrics needs to be maintained while the style should be transferred.
- No set of parallel corpora to translate between two style
- Used Encoder-Decoder sequence to sequence model and add attention to the decoder to improve the style transfer quality
- Evaluate this model on fluency, content preservation, and style



- Encoder: We use RNN with GRU cells as encoder and padded all lyrics to the same length
- Decoder: We use RNN with GRU cells as the decoder. The attention mechanism is used. An additive attention layer is added on the top of the decoder to improve the performance
- A BiLSTM classifier is used to classify lyrics as either the original style or the target style

- PPLg: general perplexity GPT-2 model is used to compute PPLg in order to reflect the fluency level in a general setting
- BLUE and BERTscore are used to evaluate the model translation
- A BiLSTM classifier to measure the percentage of correctly labeled lyrics



Results: Example of the Lyric transfer from Taylor Swift to Drake

Original	Transferred
I seen your cousin in the streets he sweet eying this booty.	I left your mark in the world whats this.
just the other day i had to shed a couple tears	i got put the greatest i had just a heavy.
and yes now im here without and see can i do	and now im blowing up and all you see is i

Results: Evaluation of our model on Fluency, Content, and Style

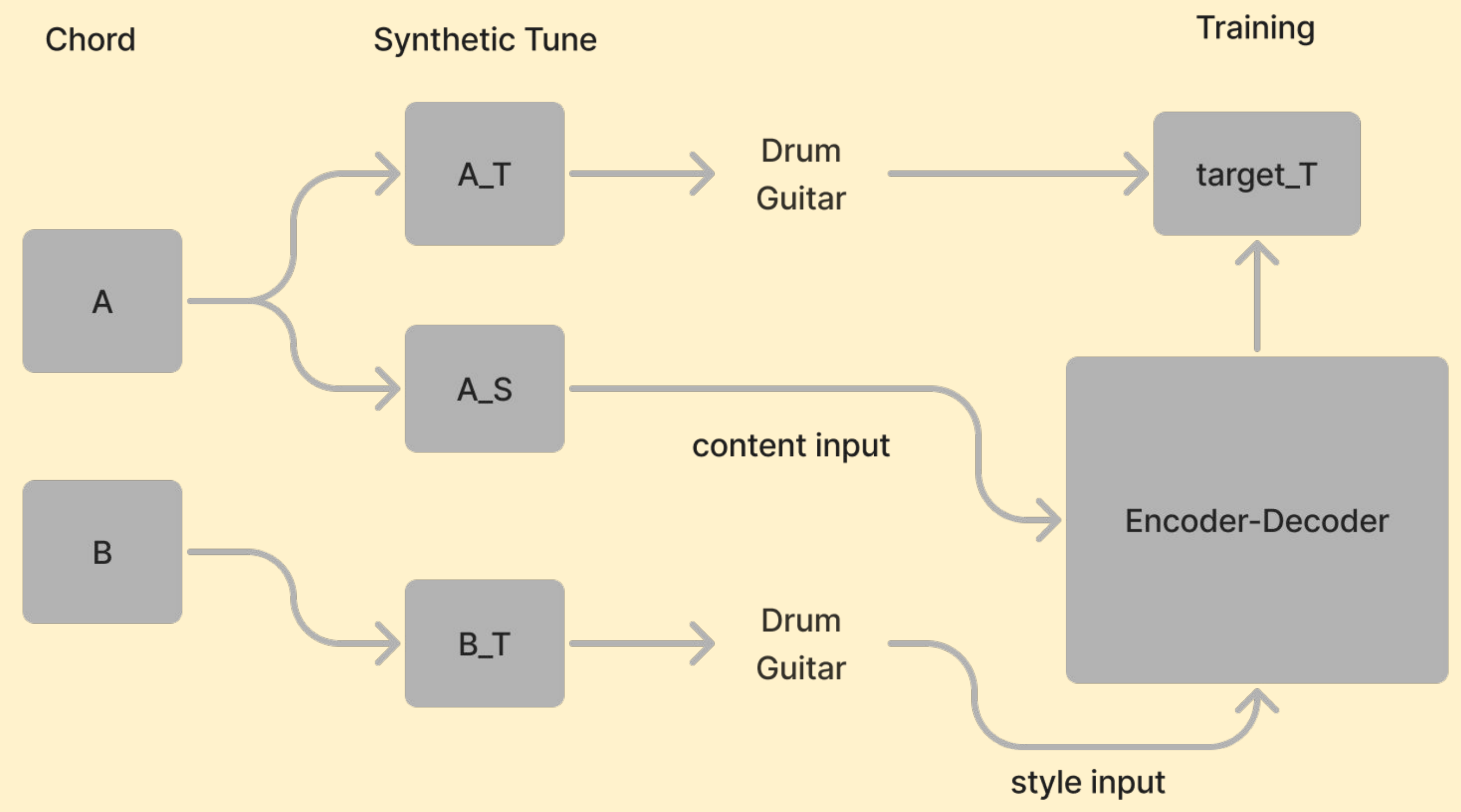
Style		
AUC real	AUC recons	AUC tsf
0.848	0.921	0.889

Content		Fluency
$BLEU$	$BERTscore$	PPL_{gpt}
65.861	0.871	26.976

Tune Style Transfer

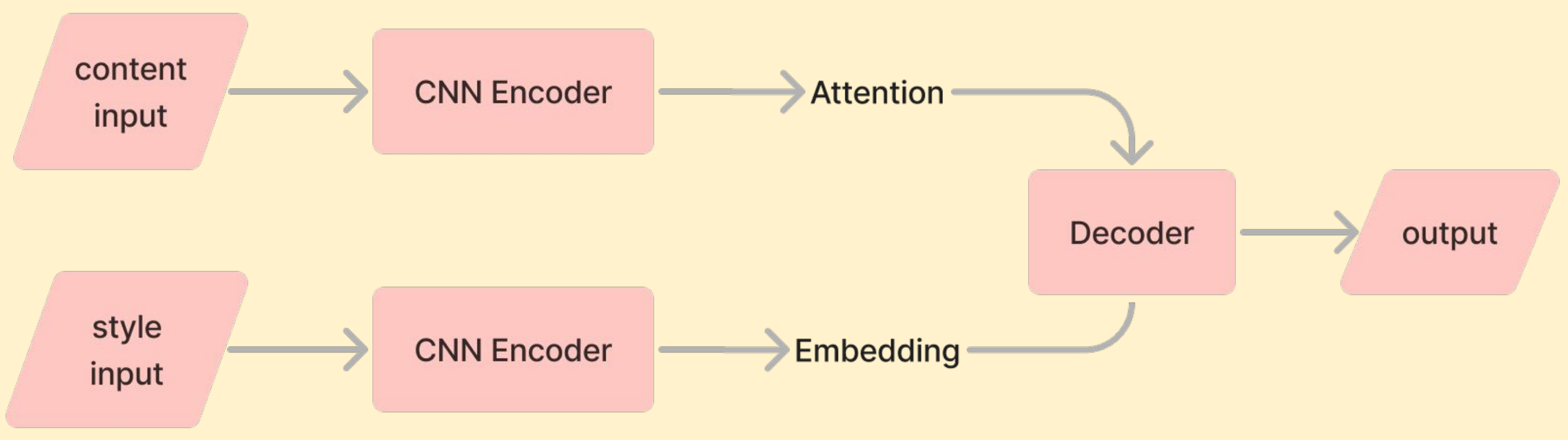
- Robust Supervised End-to-End Learning
- Synthetic Parallel Data
- Produce Any Given Combination Of Tracks

Architecture



- Starting from the chord(source A and target B), we create synthetic accompaniments in different styles (S and T)
- input: content input A_S(accompaniment for A in style S), style input B_T(a single track of B in style T)
- output: target_T(the corresponding track of the target accompaniment A_T(a single track of A in style T))

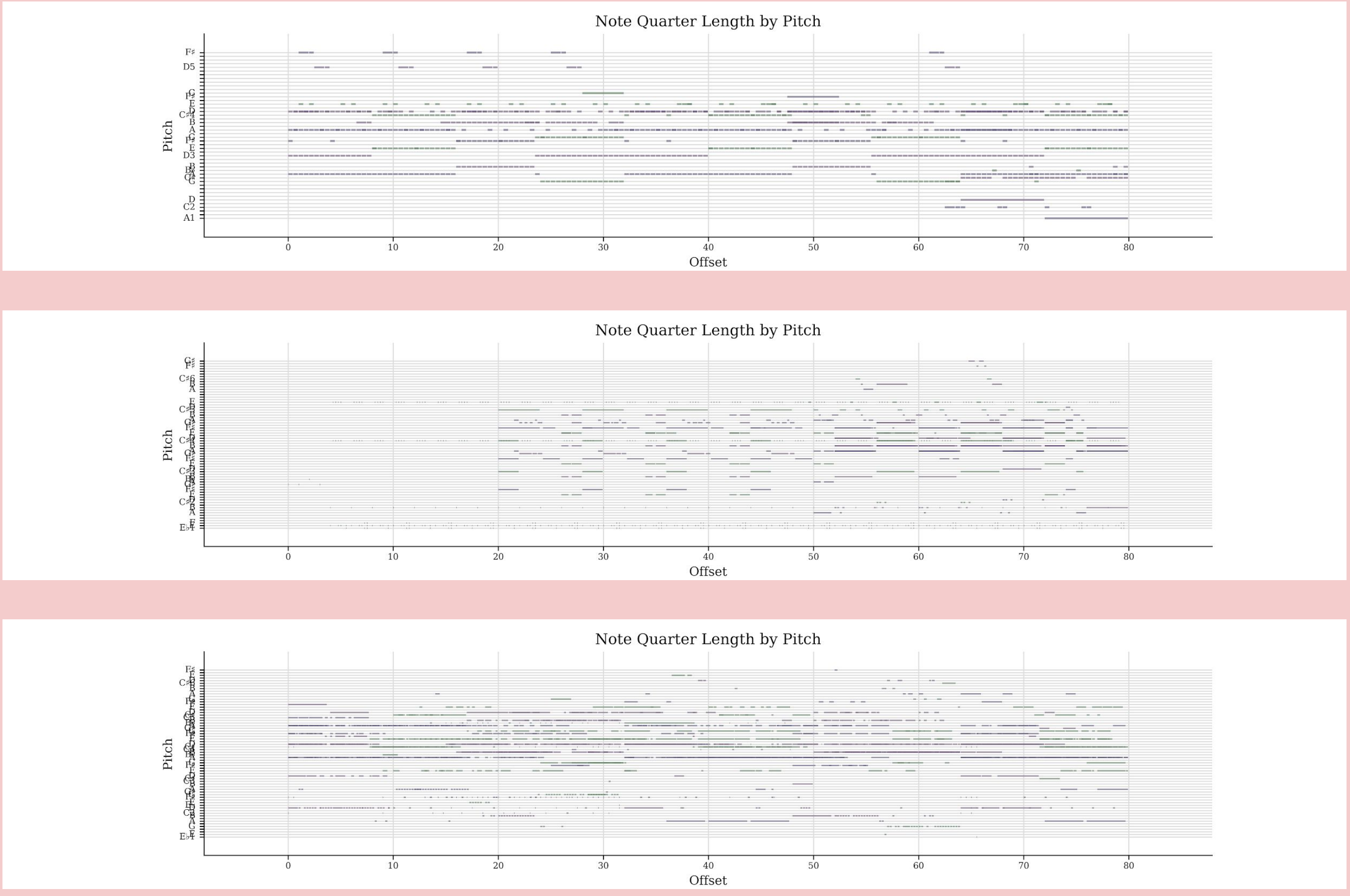
Encoder-Decoder



- Encoder: Apply CNN encoder to encode the content input and the style input respectively
- Mid-layer: Use Attention for content input to calculate the attention weight, use embedding to represent different tracks in style input
- Decoder: Combine the representations from these two encoders and use GRU to generate the corresponding output track

Results

Transfer from Taylor Swift's *Love Story* to Drake's *Find Your Love*



Conclusion

- Use encoder-decoder approach to do unsupervised style transfer and add an attention layer on the decoder to improve the performance of the transformation.
- Use bi-encoder to represent the tunes of source style and target style and apply a supervised learning approach to achieve a track-wise style transfer.

Future Work

- Maybe we can introduce the transformer architectures to further improve the performance.
- Ask humans to evaluate the lyric transfer result.
- Combine the lyrics transfer and the tune transfer to compose a complete song, and handle the alignment between lyrics and tunes.

Copyright © 2022 AI BERT