Deep Learning Techniques for Music Generation
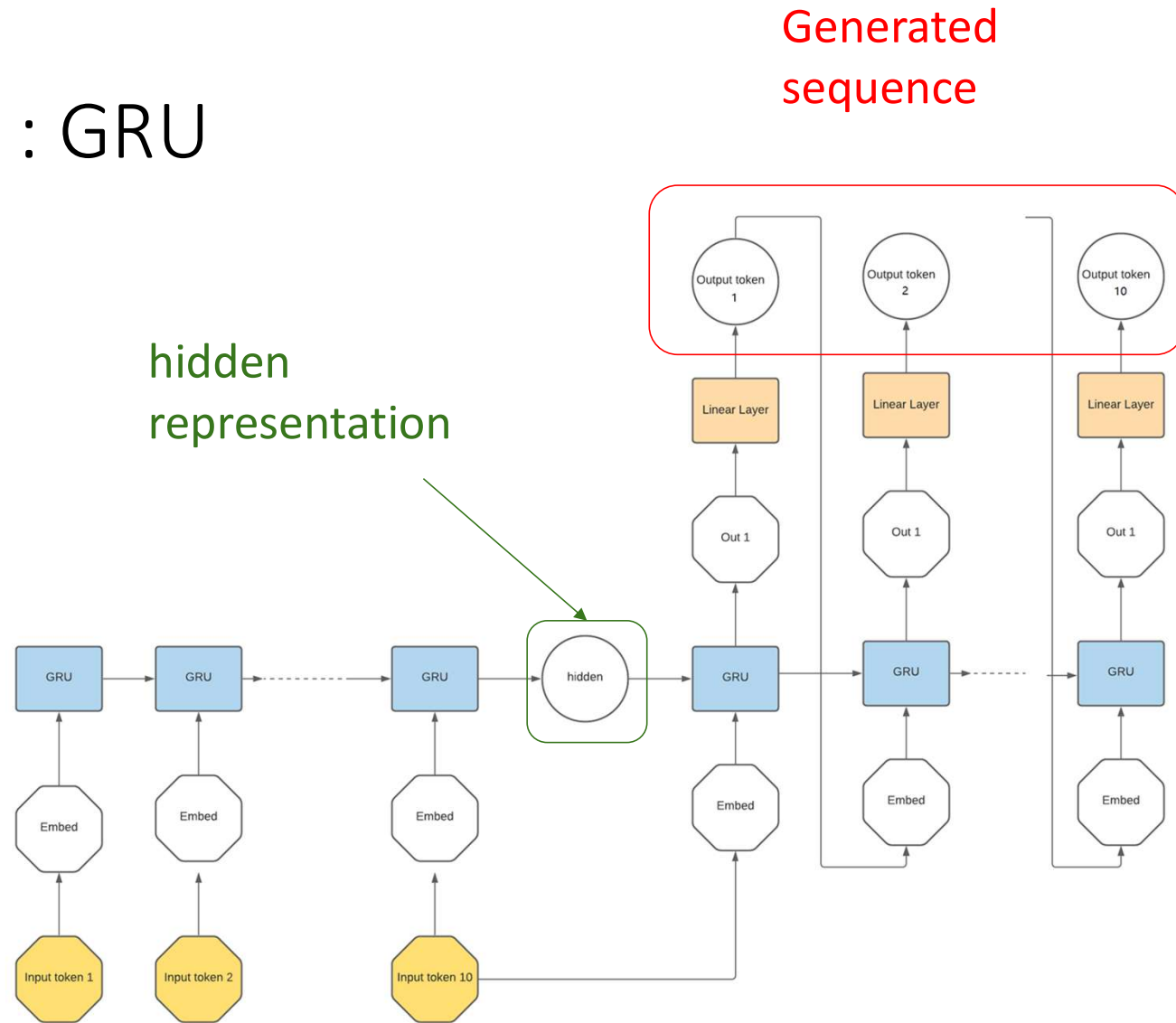
Yassine Abbahaddou
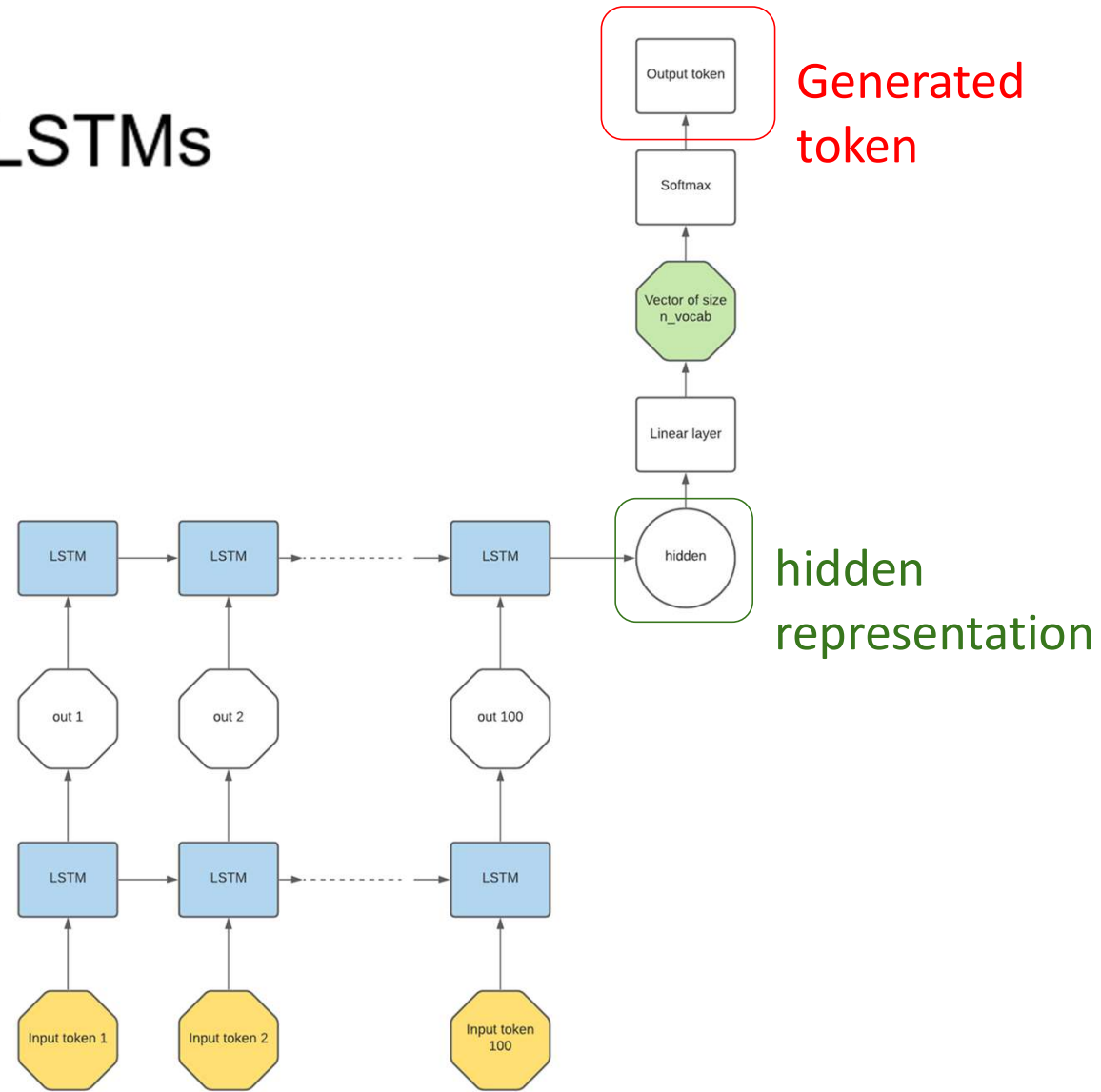
Elias Aouad

Introduction

- Goal : Generate new music using deep learning methods

- Two ways to treat this problem :
    - Use music as a raw audio
    - Encode music into textual notations

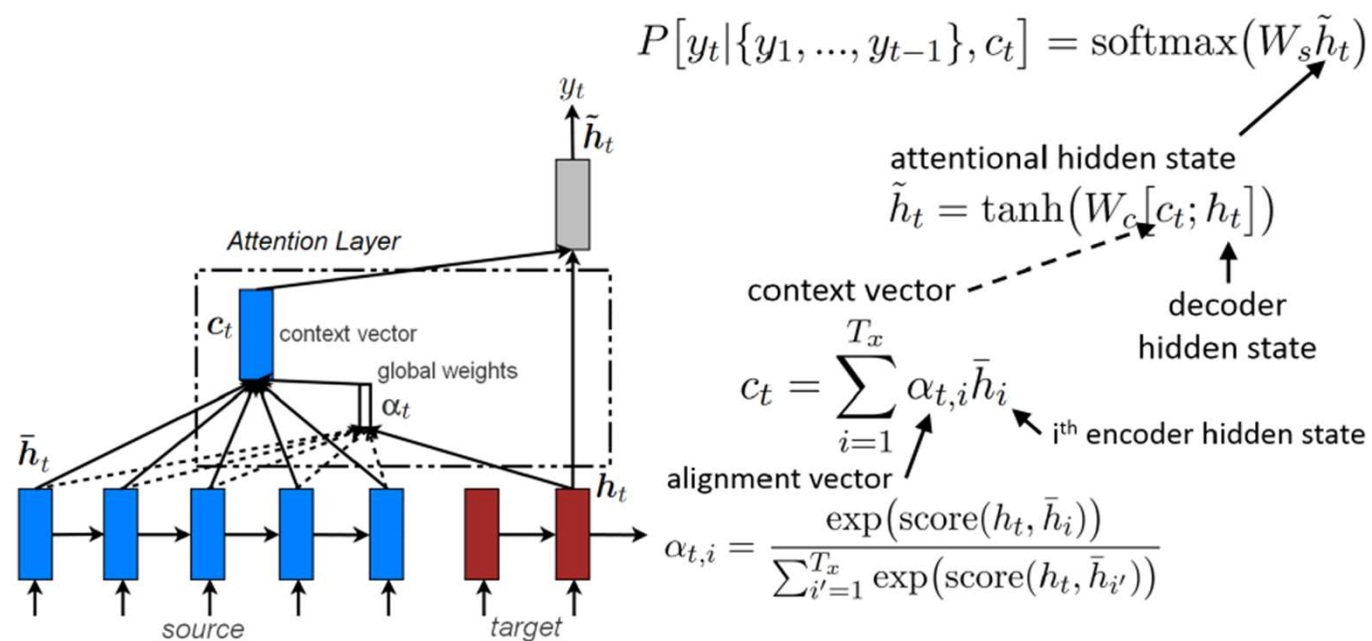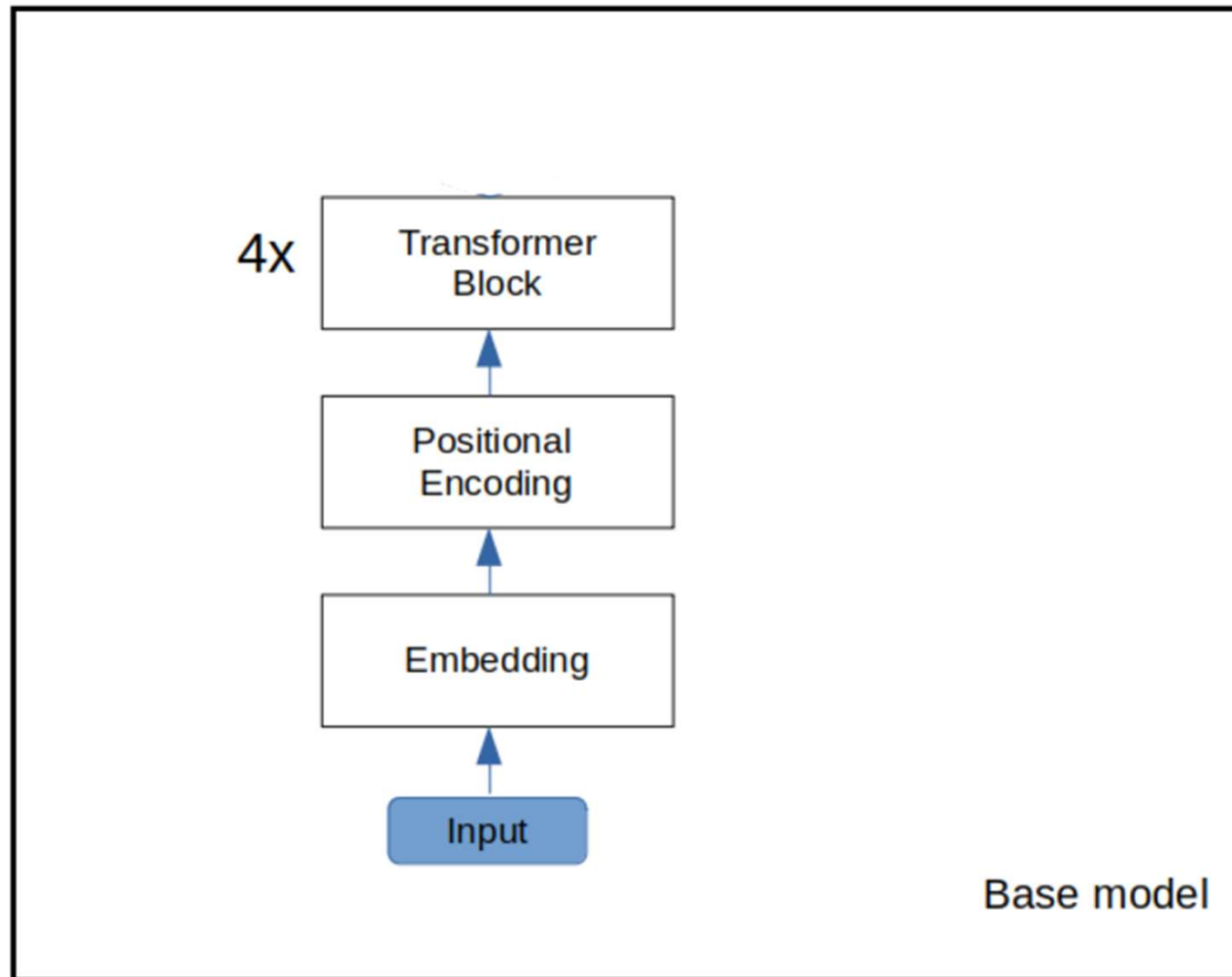- Evaluation : Use BLEU score to rate music / Discriminator error (GAN only)

Method 1 : GRU

Generated sequence

hidden representation

Method 2 : 2-layer LSTMs

Method 3 : GRU + attention mechanism

$$P\big[y_t|\{y_1, ..., y_{t-1}\}, c_t\big] = \mathrm{softmax}(W_s\tilde{h}_t)$$

attentional hidden state

$$\tilde{h}_t = \tanh\big(W_c[c_t; h_t]\big)$$

context vector

decoder
hidden state

$$c_t = \sum_{i=1}^{T_x} \alpha_{t,i}\bar{h}_i$$

ith encoder hidden state

alignment vector

$$\alpha_{t,i} = \frac{\exp\big(\mathrm{score}(h_t, \bar{h}_i)\big)}{\sum_{i'=1}^{T_x} \exp\big(\mathrm{score}(h_t, \bar{h}_{i'})\big)}$$
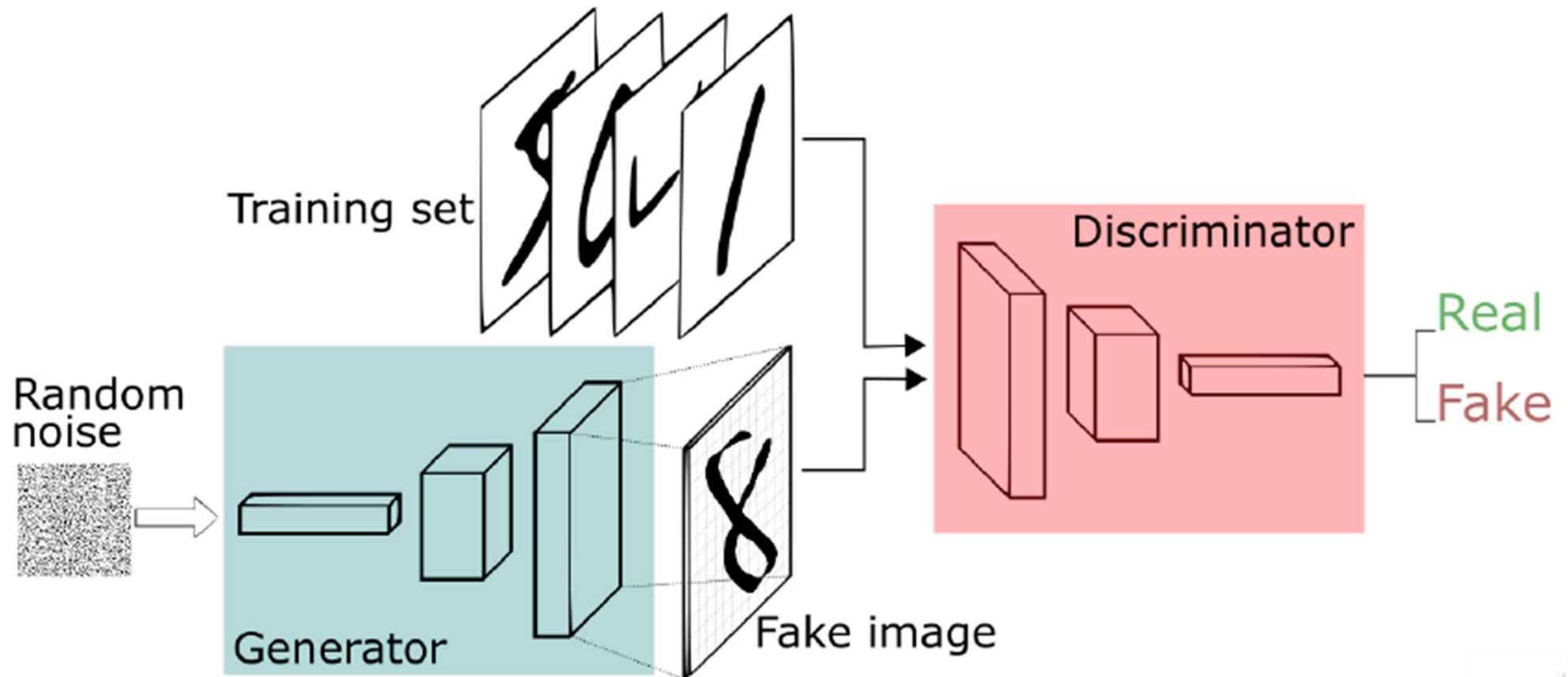
Method 4 : transformer Decoder

Method 5 ：GAN
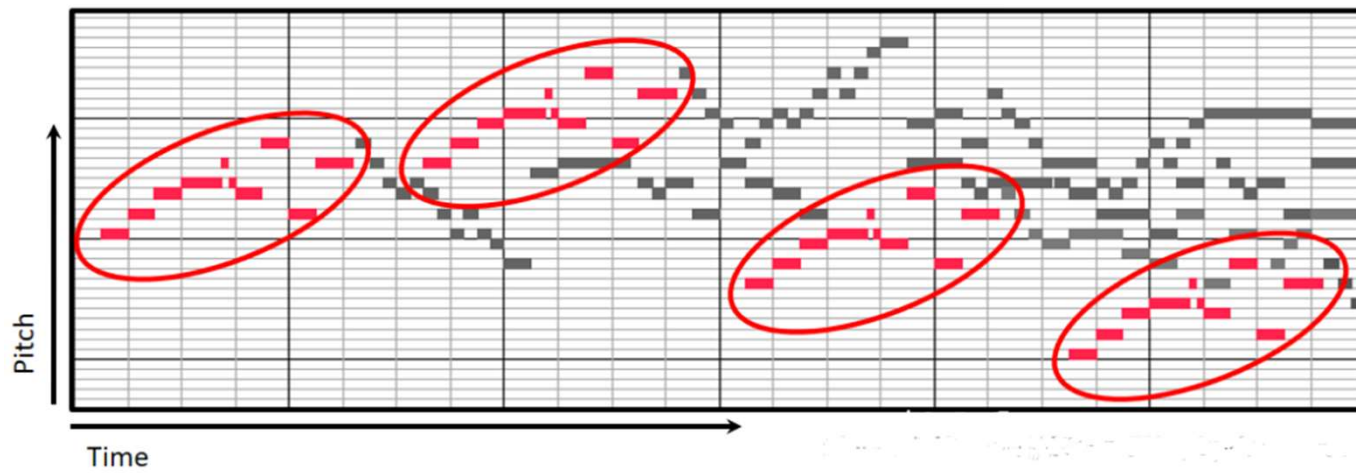
Method 5 : GAN

Metric : BLEU score

$$BLEU = \frac{\sum_{S \in C} \sum_{\text{N-gram} \in S} Count_{matched}(\text{N-gram})}{\sum_{S \in C} \sum_{\text{N-gram} \in S} Count(\text{N-gram})}$$

Results

Method	BLEU score	Discriminator	Human rating
GRU	80%	-	6.5 ($\pm$3)
2-layer LSTM	71%	-	8.5 ($\pm$1)
RNN+attention	62%	-	7.5 ($\pm$1.7)
Transformer Decoder	57%	-	9 ($\pm$1)
GAN	-	0%	9 ($\pm$0.4)

Conclusion

- We mainly relied on the latest advances in the NLP field.

- The most challenging task in music generation is the evaluation.

- In terms of results, almost all the methods produced music similar to real ones.

References

[1] R. W. Allen Huang. Deep learning for music

[2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Gen- erative Adversarial Networks. June 2014.

[3] M. Post. A call for clarity in reporting bleu scores. In WMT , 2018.

[4] A. S. B. K. Rachel Manzelli, Vijay Thakkar. Conditioning deep generative raw audio models for structured automatic music.

[5] D. Rubino. Nottingham music database.

[6] A. van den Oord et al. Wavenet: A generative model for raw audio.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. 2017.

Generated music with : transformer Decoder