

Generowanie muzyki przy pomocy uczenia maszynowego

Głównym celem pracy jest przegląd obecnie stosowanych metod i modeli uczenia maszynowego w celach generacji muzyki oraz propozycja swojego modelu, składająca się z GAN-a implementującego jako swój dyskryminator i generator architektury transformera.

TransGAN: Two Pure Transformers Can Make One Strong GAN, and That Can Scale Up

Yifan Jiang¹, Shiyu Chang^{2,3}, Zhangyang Wang¹

¹University of Texas at Austin

²UC Santa Barbara ³MIT-IBM Watson AI Lab

{yifanjiang97, atlaswang}@utexas.edu, chang87@ucsb.edu

StyleSwin: Transformer-based GAN for High-resolution Image Generation

Bowen Zhang^{1*} Shuyang Gu¹ Bo Zhang^{2†} Jianmin Bao² Dong Chen²

Fang Wen² Yong Wang¹ Baining Guo²

¹University of Science and Technology of China

²Microsoft Research Asia



Figure 1. Image samples generated by our StyleSwin on FFHQ 1024×1024 and LSUN Church 256×256 respectively.

TTS-GAN: A Transformer-based Time-Series Generative Adversarial Network

Xiaomin Li, Vangelis Metsis, Huangyingrui Wang, Anne Hee Hiong Ngu

x_l30, vmetsis, h_w91, angu@txstate.edu

Texas State University, San Marcos TX 78666, USA

STransGAN: An Empirical Study on Transformer in GANs

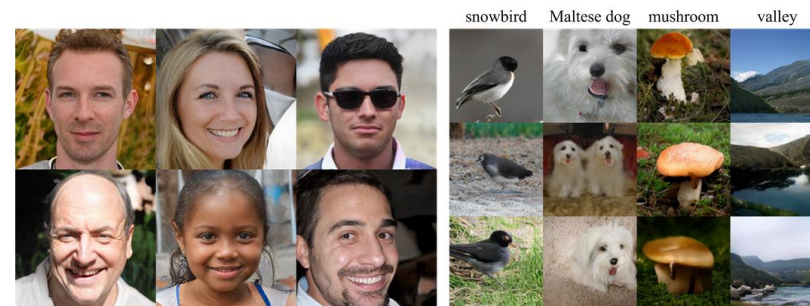
Rui Xu¹ Xiangyu Xu² Kai Chen^{3,4} Bolei Zhou¹ Chen Change Loy²

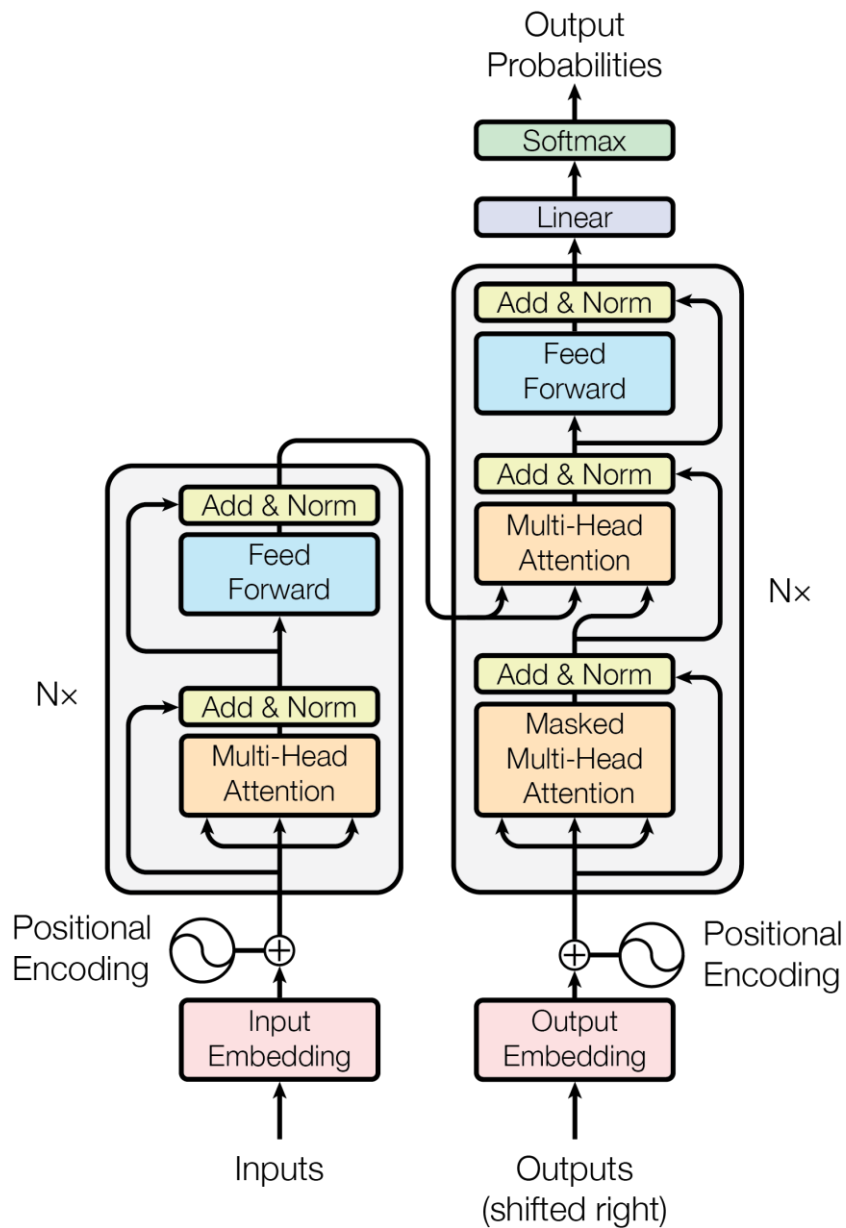
¹CUHK - SenseTime Joint Lab, The Chinese University of Hong Kong

²S-Lab, Nanyang Technological University, Singapore

³SenseTime Research ⁴Shanghai AI Laboratory

Abstract

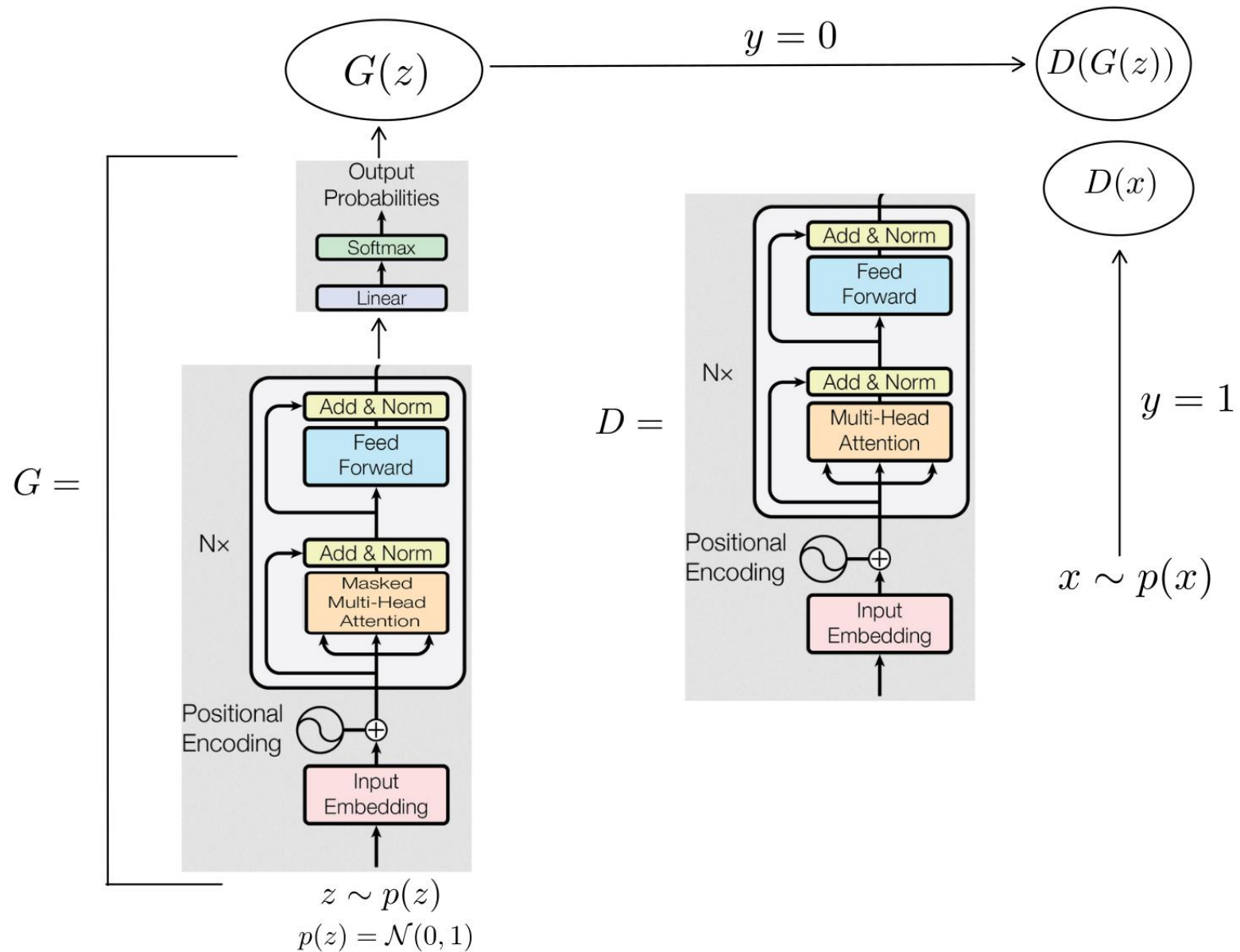




Architektura transformera:

- enkoder + dekodery
- „kontekst” z enkodera jest przekazywany do dekodera, który bierze go pod uwagę podczas generacji kolejnych tokenów
- mój przypadek nie wymaga analizy kontekstu przez co enkoder może zostać usunięty (razem z drugą warstwą Multi-Head Attention)

Architektura modelu



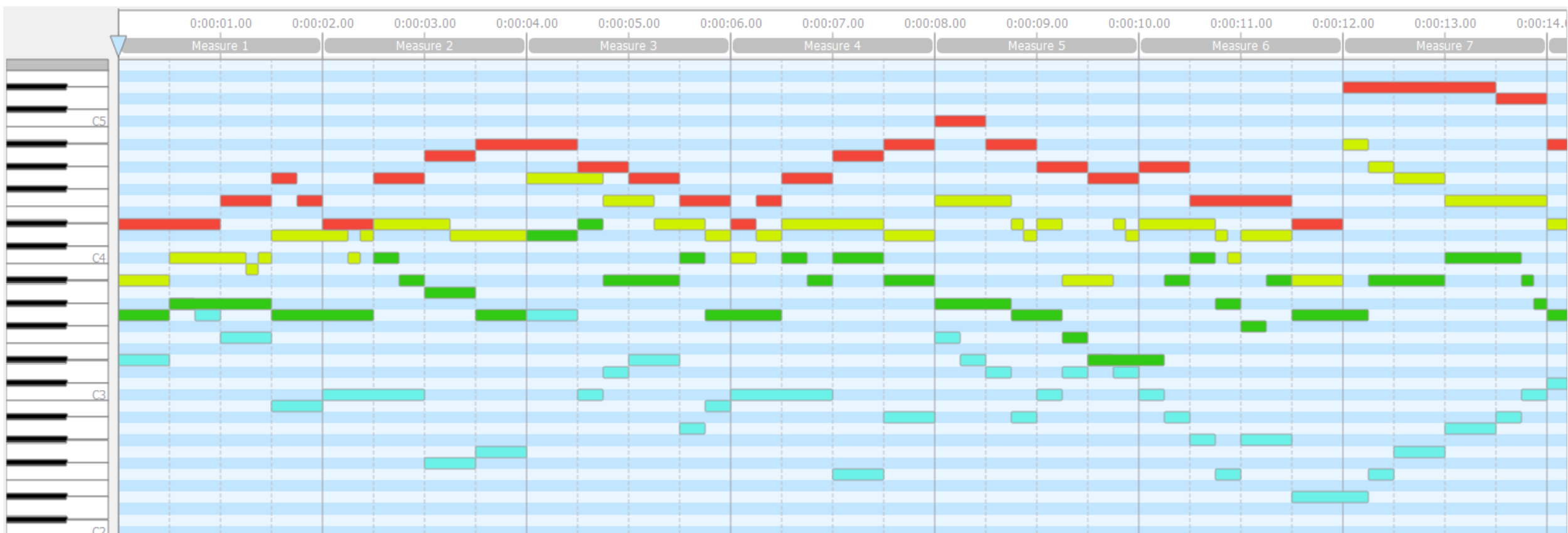
Celem treningu jest dojście do momentu w którym generator dostarcza tak dobre dane, że dyskryminator nie jest ich w stanie odróżnić od prawdziwych przez co jego wynik daje zawsze 0.5

Matematycznie możemy zapisać:

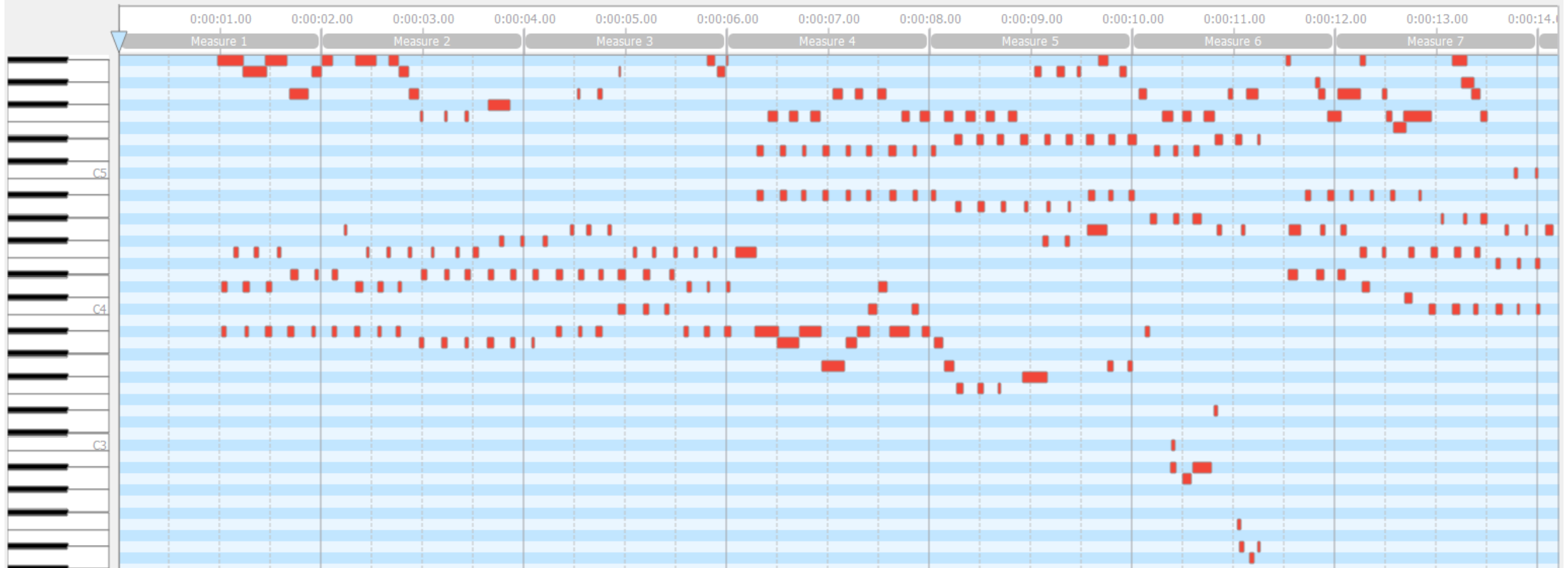
$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

Zbiory danych używane w pracy

Bach chorales



Maestro



Technologie



Biblioteka pretty_midi