



Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação



IA376L - Deep Learning Aplicado a Síntese de Sinais
Prof^a. Dr^a. Paula Dornhofer Paro Costa

Síntese de timbres de instrumentos musicais através de Redes Adversárias Generativas

Alunos: Gabriel Santos Martins Dias. RA: 172441.
Gleyson Roberto do Nascimento. RA: 043801.
Patrick Carvalho Tavares R. Ferreira. RA: 175480.

Abril/2022

Proposta:

Neste projeto, temos por objetivo a síntese generativa dos timbres de instrumentos musicais como flauta, violão, piano, órgão, metais (sopro), cordas e demais exemplares que constam no Nsynth Dataset através do uso de Redes Adversárias Generativas.

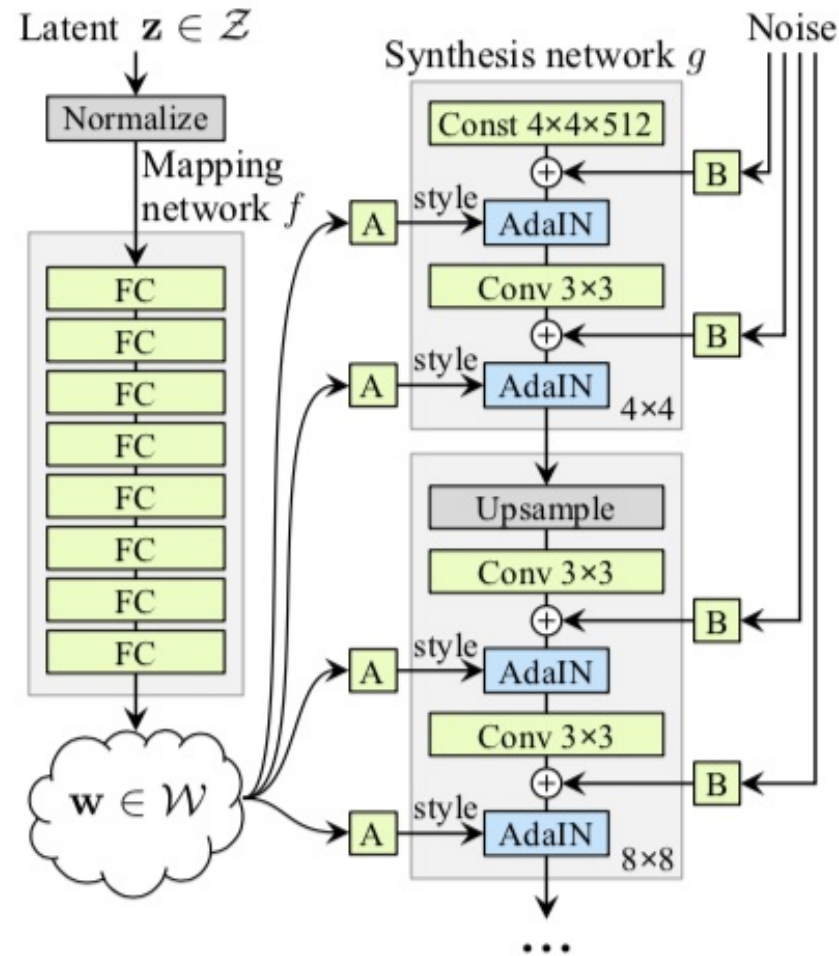
Datasets (Active Loop - HUB): nsynth train, nsynth test, nsynth validation

The screenshot displays the ActiveLoop HUB interface for the **nsynth-train** dataset. The left sidebar contains navigation icons for Datasets and Docs. The main content area shows the dataset details under the 'DETAILS' tab, including a search bar, a 'Query the dataset' button, and a 'main' dropdown menu. The dataset name 'nsynth-train' is prominently displayed, along with a code snippet: `ds = hub.load('hub://activeloop/nsynth-train')`. Below this, metadata is provided: Created by amalmathewomdena, Last updated: 02/06/2022, Size of dataset: Not Available, and Number of tensors: Not Available. A 'README' section indicates that the readme is currently empty.

On the right side, a grid of audio waveform visualizations is shown, each representing a different audio sample. A modal window titled 'Index: 140150' is open, displaying a list of tensors available for the selected index. The tensors listed are:

- qualities
- instrument_family
- instrument_source
- audios
- note
- instrument
- pitch
- velocity
- sample_rate

Arquitetura da Rede Neural: GANSynth



Assim, o modelo amostra um vetor aleatório z de uma gaussiana esférica e o executa através de uma pilha de convoluções transpostas para aumentar a amostragem e gerar dados de saída $x = G(z)$.

Ferramentas:

Ferramenta	Descrição
Google Colab	Ferramenta para elaboração dos notebooks e códigos em linguagem Python 3.8
Active Loop - HUB	Ferramenta para manipular o repositório com o Nsynth Dataset
Tensorflow	Ferramenta principal de manipulação de tensores e construção da GAN
Pytorch	Ferramenta secundária de manipulação de tensores e rede neural
Magenta	Ferramenta para manipulação de áudio em inteligência artificial
Librosa	Ferramenta de análise de <i>features</i> de áudio
SDV – Table Evaluate	Ferramenta para avaliação de dados tabulares
Seaborn	Ferramenta para <i>Data Visualization</i>

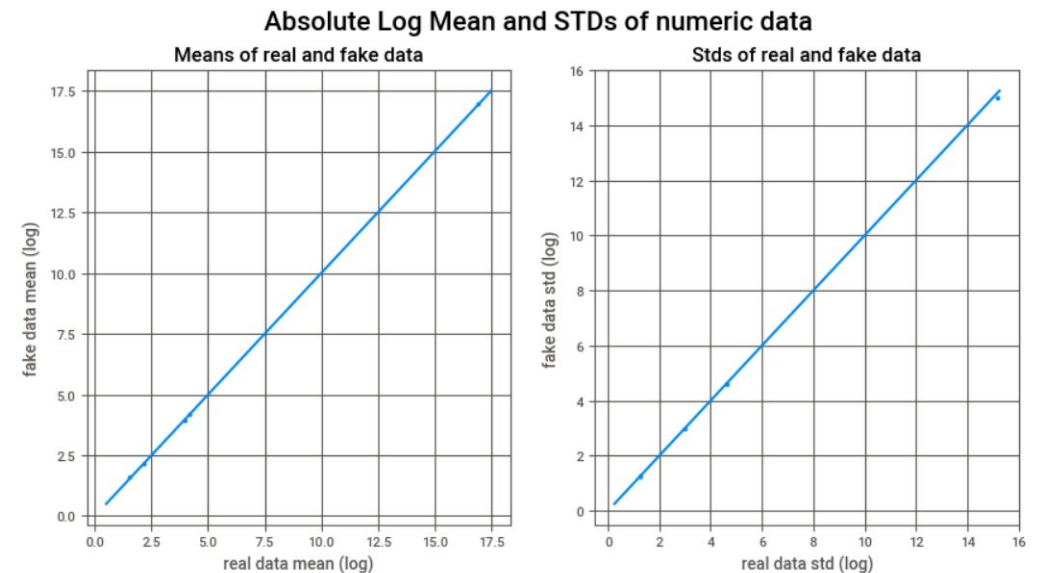
Avaliação I:

Quantitativa:

1) Teste Kolmogorov–Smirnov para avaliar o score entre o áudio real e sintetizado;

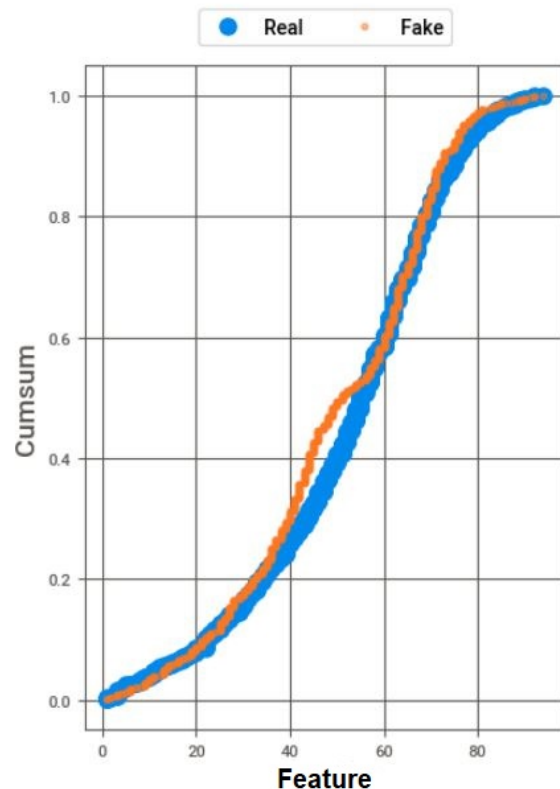
Qualitativa:

1) Gráfico log-log da média dos dados reais x média dos dados sintetizados; gráfico log-log do desvio-padrão dos dados reais x desvio-padrão dos dados sintetizados (para verificar tendência de 45°);



Qualitativa:

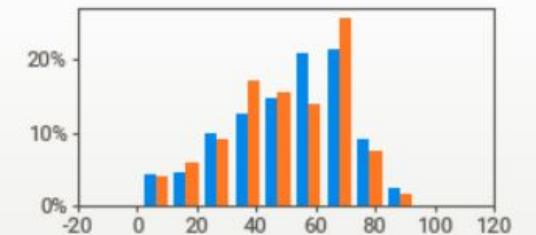
2) gráfico da soma cumulativa dos dados reais e sintetizados (para verificar diferença de distribuição);



3) histograma dos dados reais e sintetizados (para verificar diferença de distribuição);

MAX	104	94
95%	79	77
Q3	66	67
MEDIAN	55	51
AVG	51	50
Q1	38	36
5%	14	13
MIN	1	1

RANGE	103	93.0
IQR	28.0	31.0
STD	20.0	19.9
VAR	398	396
KURT.	-0.414	-0.658
SKEW	-0.473	-0.367
SUM	4.0M	50,175



Music Information Retrieval: através das análises espectrais existentes na Biblioteca Librosa, fazer o comparativo entre as *features* espectrais do áudio real e do áudio sintetizado. Serão analisadas as seguintes *features* espectrais:

STFT, espectrograma, mel espectrograma, cromagrama, MFCC, RMS, spectral *bandwidth*, spectral *rolloff*, spectral *flatness* e *zero-crossing rate*.

Referências Bibliográficas:

[1] Engel, J.; Agrawal, K. K.; Chen, S.; Gulrajani, I.; Donahue, C.; Roberts, A.; "**Gansynth: Adversarial Neural Audio Synthesis**"; ICLR; 2019;

Nesta referência temos a apresentação da GANSynth e sua arquitetura, de forma que representa a principal referência para este projeto;

[2] Karras, T.; Laine, S.; Aila, T.; "**A style-based generator architecture for generative adversarial networks**"; CoRR; abs/1812.04948; 2018b;

Nesta referência temos a apresentação da abordagem da síntese de imagens que inspirou o artigo da GANSynth, de forma que ela apresenta maiores detalhes com relação ao processo de síntese da GAN;

[3] Muller, M.; "**Information Retrieval for Music and Motion**"; 1ª Edição; Editora Springer; 2010;

Nesta referência temos as definições de *Music Information Retrieval* (MIR) e, conseqüentemente, das *features* a serem observadas para a comparação entre os dados reais e os dados sintetizados.

Resultados Esperados:

A síntese de áudio ainda está em franca evolução, todavia, pelos artigos envolvendo o uso da GANSynth, o que esperamos é que a diferença entre os dados sintetizados e os reais não seja tão grande tanto nas análises quantitativas quanto qualitativas e, principalmente, que respeite a ordem de grandeza das *features* espectrais de MIR.