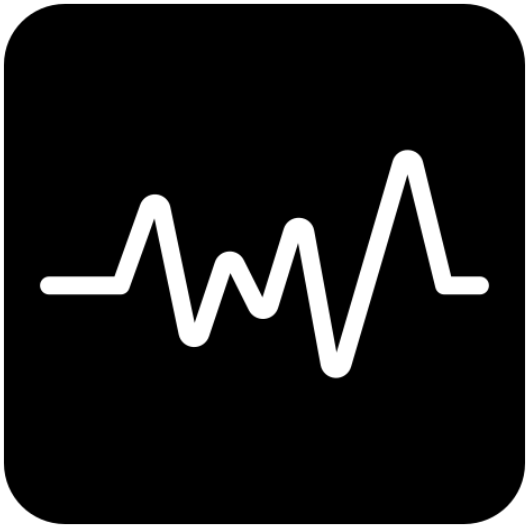
Universitat Pompeu Fabra Barcelona

Real-time Neural Texture Sound Synthesis with physically-driven controls using synthetic-to-real unsupervised Domain Adaptation

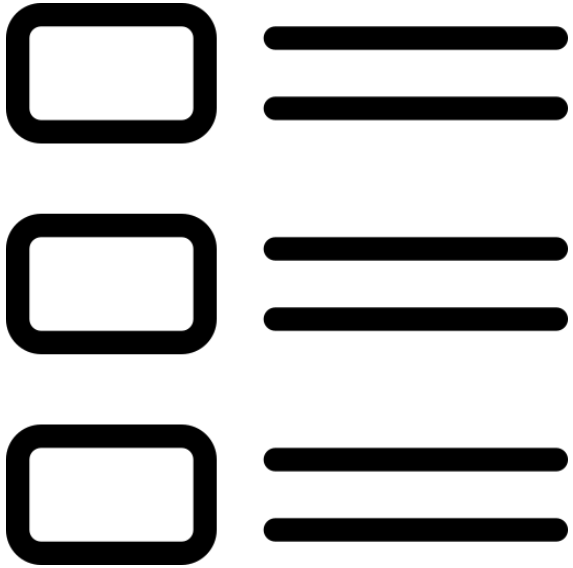


Sound and Music Computing Master - UPF

2022/2023

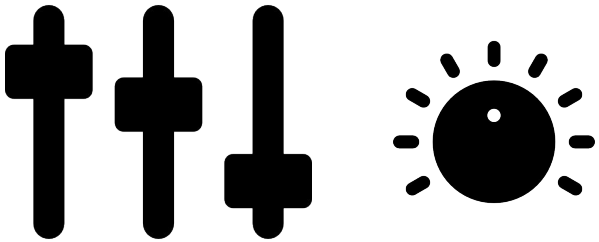
Matteo Fabbri
Lonce Wyse

INDEX



INTRODUCTION

Continuous controls

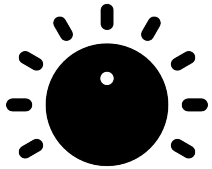


Environmental sounds

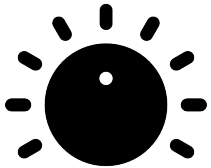
INTRODUCTION

Physically driven

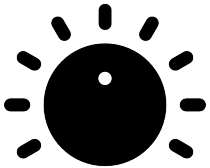
Synthesis Control Parameters



Average bubbles rate

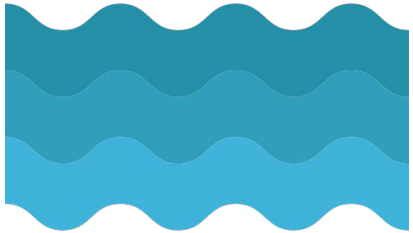


Minimum bubbles size



Maximum bubbles size

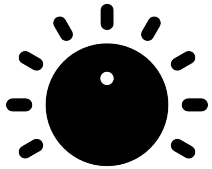
Environmental sounds



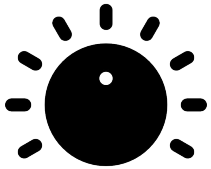
CHALLENGES

Physically driven

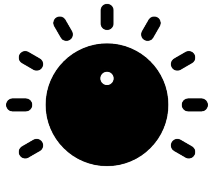
Synthesis Control Parameters



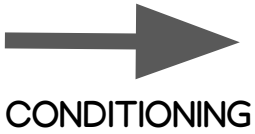
Average bubbles rate



Minimum bubbles size



Maximum bubbles size



CONDITIONING

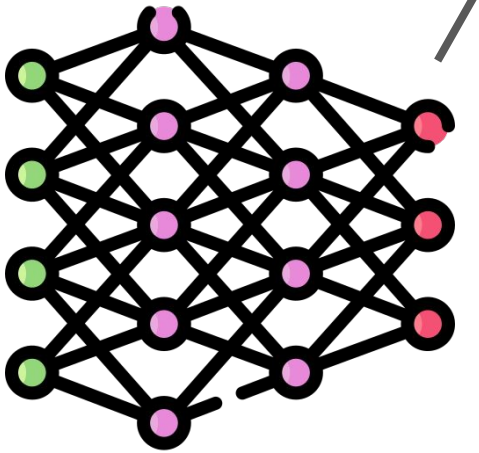
DATASET

Neural Network

Synthesiser

GROUND TRUTH

LOSS



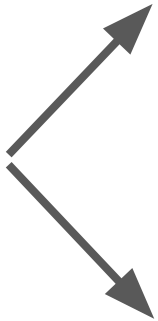
STATE OF THE ART

- DDSP – Differentiable Signal Processing (Esling et al., 2020)
- Automatic characterization and generation of music loops and instrument samples for electronic music production (Ramires, A., 2023)
- Huzaifah Bin Md Shahrin, M. (2020, December 21). Directed Audio Texture synthesis with Deep Learning. NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore
- Sound Model Factory (Wyse, 2020)

APPROACH and METHODOLOGIES

Abstract and modular Software Framework exposing a JSON dictionary interface

```
{
synthesis_Controls_Names : list()
number_Conv_Layers : int()
input_Transforms : torchaudio.Transform()
etc…
}
```



Specific Synthesisers instances

- Pre-determined Synthesis Control Parameters
- Specific sound types and Convolutional NN Architectures

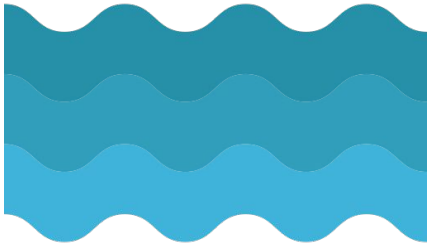
Waterflow sounds

```
synthesis_Controls_Names :
['avg_BubblesRate', 'minimum_BubblesSize']
number_Conv_Layers : 4
input_Transforms : [Resample(), MelSpectrogram()]
etc…
```

Fire sounds

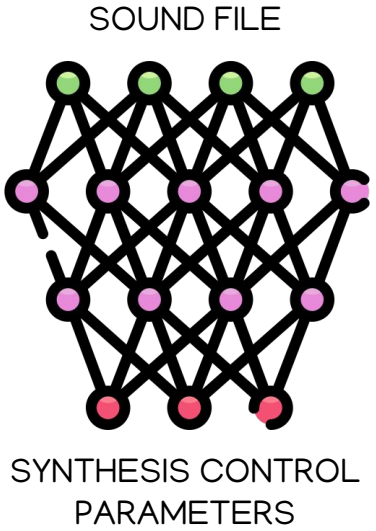
```
synthesis_Controls_Names :
['crackling_Amount', 'hissing_Amount']
number_Conv_Layers : 6
input_Transforms : [Resample()]
etc…
```

APPROACH and METHODOLOGIES



Labelled
Synthetic
Audio dataset
Data augmentation

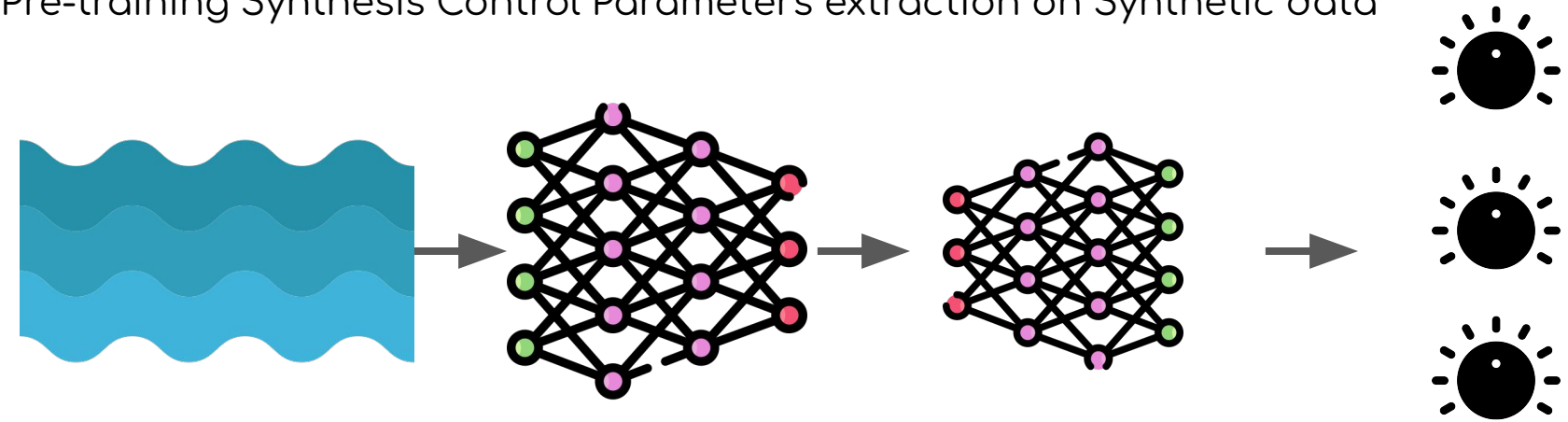
Unlabelled Real
Audio dataset
Segmented subset



Synthesis Control
Parameters
extractor

APPROACH and METHODOLOGIES

Pre-training Synthesis Control Parameters extraction on Synthetic data



Synthetic Audio dataset with labels

Source Domain CNN

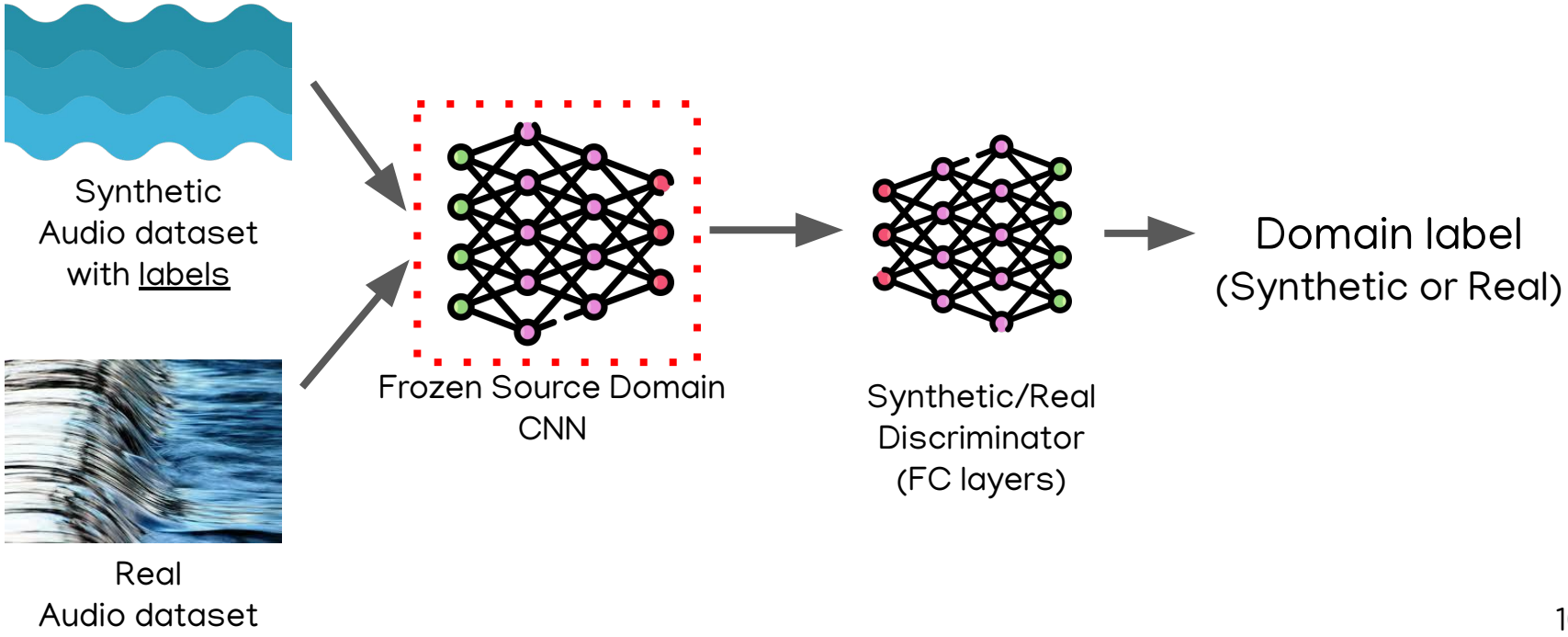
Regressor FC layers

Synthesis Control Parameters

LOSS AGAINST GROUND TRUTH

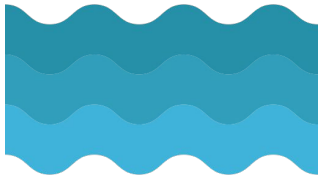
APPROACH and METHODOLOGIES

Domain adaptation - Domain Classifier training



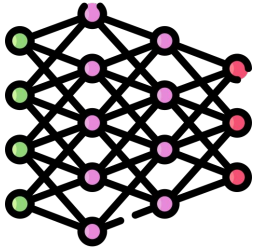
APPROACH and METHODOLOGIES

Domain adaptation - Convolutional layers adaptation

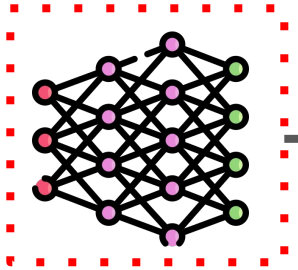


Synthetic
Audio dataset
with labels

Real
Audio dataset



Target Domain CNN

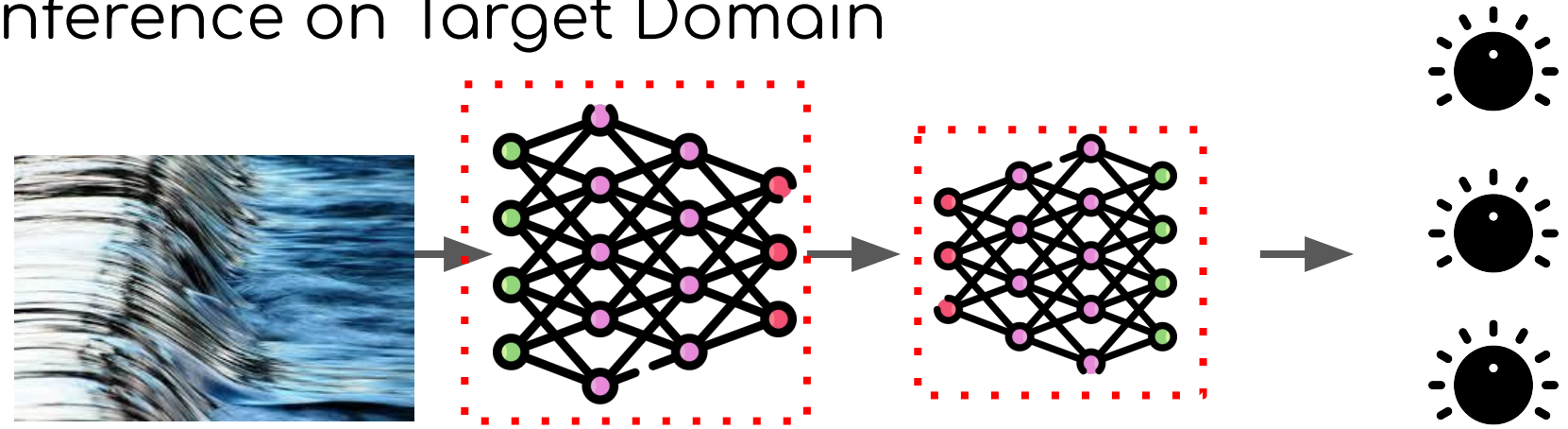


Frozen
Synthetic/Real
Discriminator
(FC layers)

Domain label
Target: Synthetic

APPROACH and METHODOLOGIES

Inference on Target Domain



Real
Audio dataset

Target Domain
CNN

Regressor
FC layers

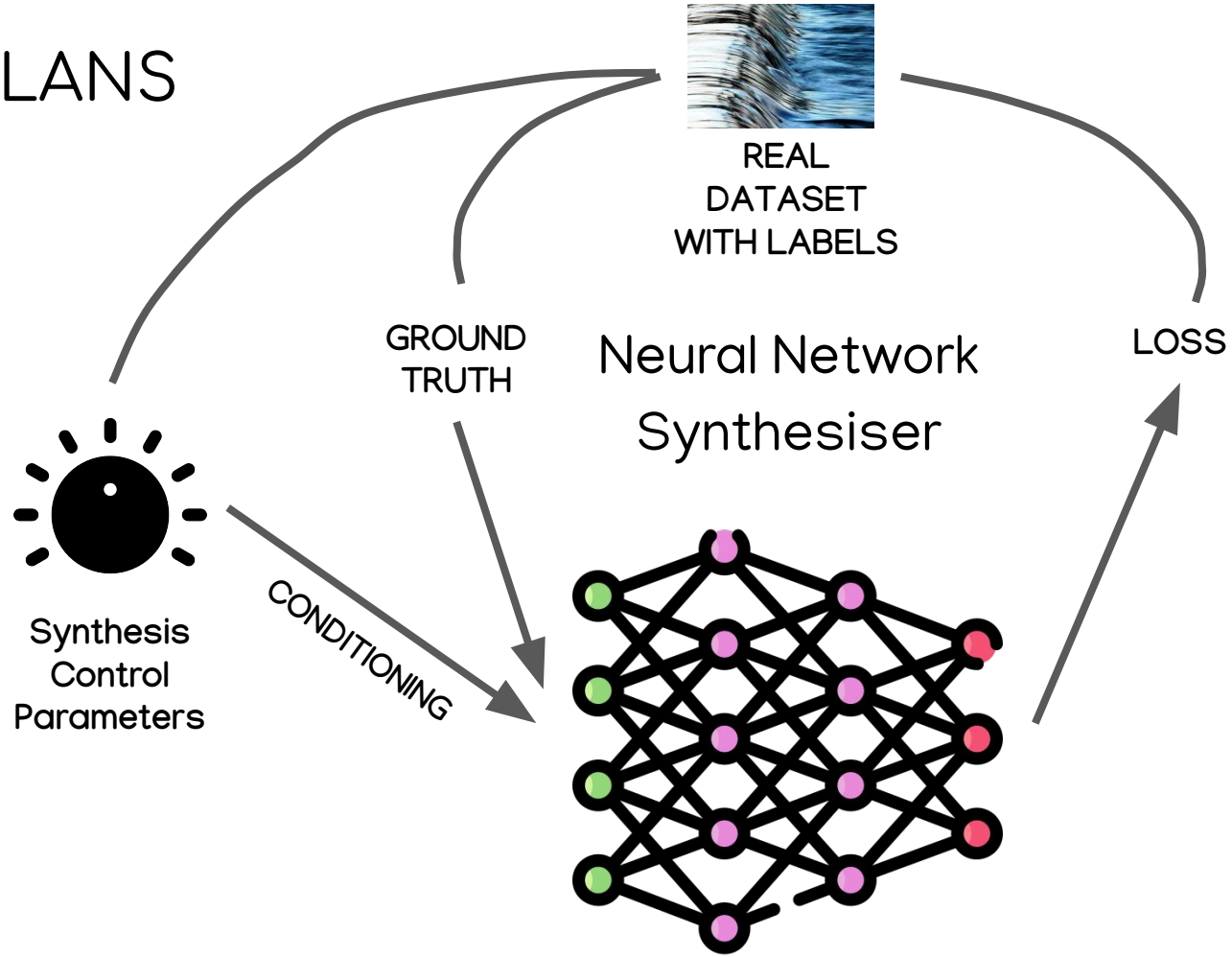
Synthesis Control
Parameters

EXPERIMENTS and RESULTS

- Non-**normalized data** produces NN modal collapse
- **Perceptually-relevant** Audio representations improve NN performance
- **Domain Adaptation** improves NN performance on Real data

FUTURE PLANS

Synthesis



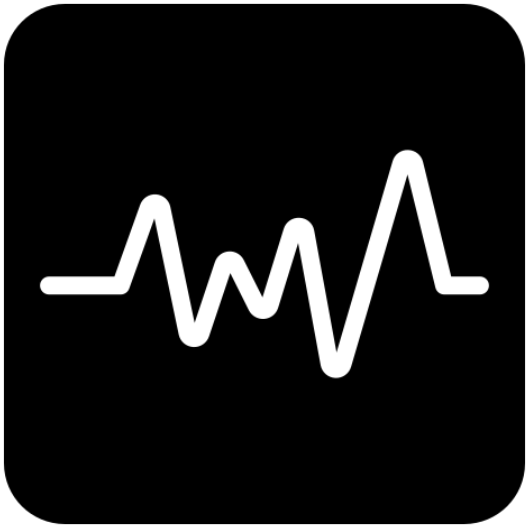
FUTURE PLANS

Synthetic dataset generation

- Evaluate the **CORRELATION between real and synthetic datasets**
  - HOW TO ANALYSE AND QUANTIFY THE DISTRIBUTION OF SYNTHESIS CONTROL PARAMETERS IN REAL DATA ?

THANK YOU FOR YOUR ATTENTION!

Any questions?



Have a look at my repo:
https://github.com/Metiu-Metiu/SMC_thesis
Leave a star if you can !!