A MULTI-OBJECTIVE APPROACH FOR SUSTAINABLE DEEP LEARNING

Constance Douwes[1] and Philippe Esling[1]

[1] IRCAM CNRS – UMR 9912, 1 Place Igor Stravinsky, F-75004 Paris, France

Contact : douwes@ircam.fr

Abstract

In recent years, the deep learning community has largely focused on the accuracy of deep generative models resulting in impressive improvements in several research fields. However, this scientific race for accuracy comes at a tremendous computational cost, which incurs vast **energy consumption** and greenhouse gas emissions. At the heart of this problem are the measures that we use as a scientific community to evaluate our work. We introduce here the idea of relying on a **multi-objective measure** based on Pareto optimality, which simultaneously integrates the models accuracy, as well as the **environmental impact** of their training and inference.

Deep accuracy madness

Deep learning holds most state-of-the-arts in various tasks.
However some problems of deep learning :

- Networks can have up to **billions of parameters**
- **Gains in accuracy** now appear always **linked to increased size**
- Extremely demanding in **computation**, **energy and memory**

Example of GPT-3 model (NLP)

- **175 billon parameters** where training takes 355 years on a V100 GPU
- Carbon footprint for training equivalent to **driving to the moon and back**

Generative models for audio

Generative models are a flourishing class of deep learning approaches

- Deal to **generate novel data** based on existing examples.
- Common models are auto-regressive, VAEs, GANs, normalizing flows

When applied to audio generation, models are generally heavy:

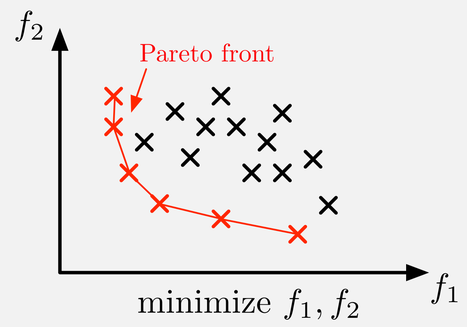
- Waveform is a **high dimensionality** vector
- Strong **temporal dependencies** at different scales

Pareto optimality

Optimization problems involving **conflicting objectives** to be optimized simultaneously :

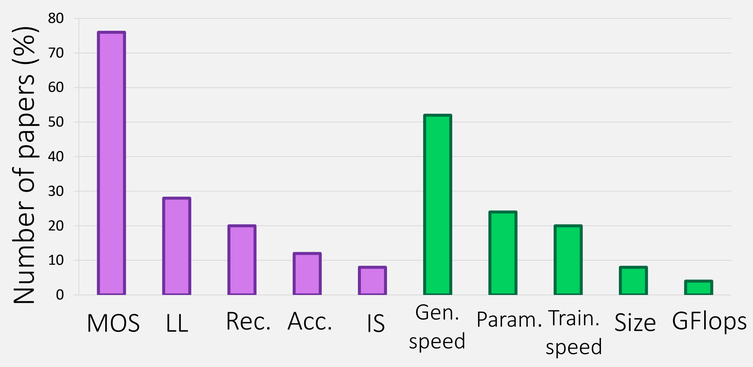
$$\min_{x \in X} (f_1(x), f_2(x), \ldots, f_k(x))$$

The set of all these pareto optimal solution is called the ***Pareto front.***



Experiments

- We reviewed all state-of-the-art audio generative models (2016 - 2020)
- We study the metrics for the evaluation and comparison :



MOS	Mean Opinion Score
LL	Log-Likelihood
Rec.	Reconstruction
Acc.	Accuracy
IS	Inception Score

Gen. speed	Generation speed
Param.	Number of parameters
Train. speed	Training speed
Size	Memory size
Gflops	Gigaflops

Difficulties to find a common metric to evaluate models

- Most of used metrics are on accuracy either than performance
- No energy-based criterion
- Best *trade-off* : **accuracy** or **energy efficiency?**

We selected models with sufficient training : hardware and training time

We create a Pareto front with :

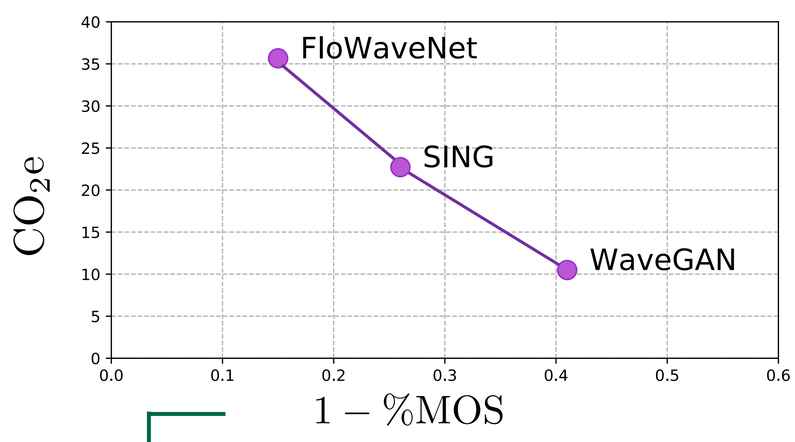
f_1 : environmental cost (either for training or for inference)
f_2 : accuracy of the model

Training cost

Carbon emissions estimation (in $kgCO_2eq$) per training can be expressed as : $\mathrm{CO_2e} = \alpha \times n \times p_{max} \times t$

α	Average electricity emission factor
p_{max}	Maximum Power of the GPU
n	Number of GPUs
t	Training time

We took $\alpha = 0,437 \ \mathrm{kgCO_2eq/kWh}$ (2018 global average)



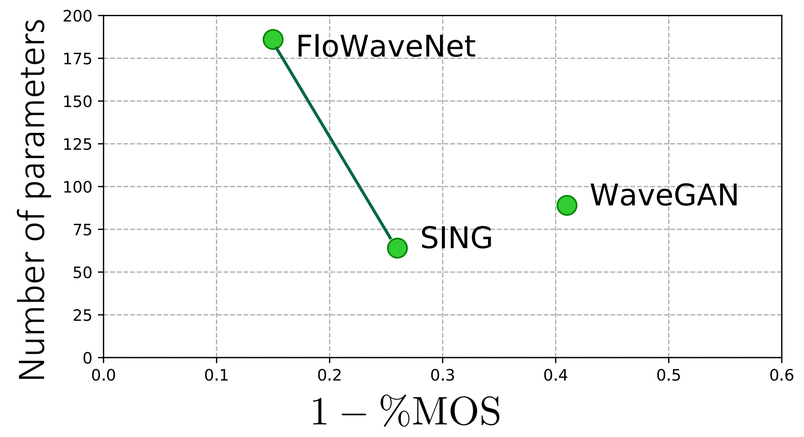
Accuracy score

We rely on the **MOS** as it is the most popular measure among our surveyed papers. This score is highly dependent on each experimental setup, we compute : $\%\mathrm{MOS} = \frac{\mathrm{MOS}(model)}{\mathrm{MOS}(truth)}$

The goal is to maximize this ratio, and thus to minimize $1 - \%\mathrm{MOS}$

Inference cost

The **number of parameters** of a model is straightforward and correlates with the number of operation of a model, and thus the power consumption of any device used to run the model.



→ The three models are Pareto optimal in training, whereas WaveGan is **sub-optimal** in inference.

Conclusions

- The lack of training details affected our work : authors must report the training time & hardware or use online tool[1] to report actual CO_2
- Models that are sub-optimal should be **discredited**
- Our approach is **generic**, and could be applied to any type of model or input data

[1] https://mlco2.github.io/impact/

University at Buffalo The State University of New York

ircam Centre Pompidou · cnrs · SORBONNE UNIVERSITÉ · ACIDS · SIF Société Informatique de France