Performance And Energy Balance: A Comprehensive Study Of State-Of-The-Art Sound Event Detection Systems

Link to paper ↓

Francesca Ronchini[1], Romain Serizel[2]

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
[2]Universite de Lorraine, CNRS, Inria, Loria, Nancy, France

Overall goal

- Study the correlation between Sound Event Detection (SED) systems performance metrics and energy consumption-related measures.
- Provide insights for a better balance between performance and energy efficiency in SED systems development.

Why studying energy consumption?

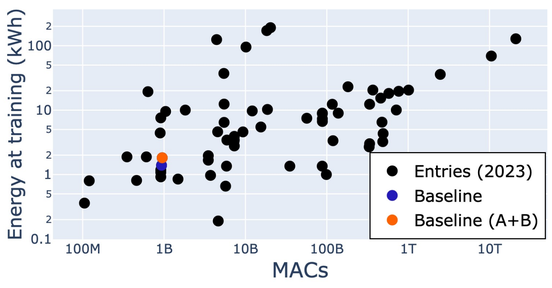
- The environmental impact of deep learning models has been dominated by the persistent demand for high accuracy and effectiveness.
- Comparing accurately the energy of different models is not straightforward.
- There has been an increasing trend of models parameter complexity.

Analysis setup and evaluation metrics

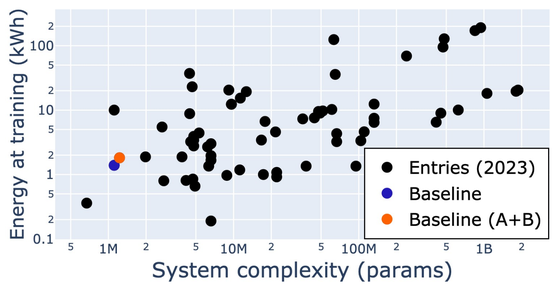
- Focusing on 2023 submissions with normalized energy consumption.
- Performance evaluation with polyphonic sound detection score (PSDS).
- Energy weighted polyphonic sound detection score (EW-PSDS):

$$\text{EW} - \text{PSDS} = \text{PSDS} * \frac{\text{kWh}_{\text{baseline}}}{\text{kWh}_{\text{submission}}} \quad (1)$$

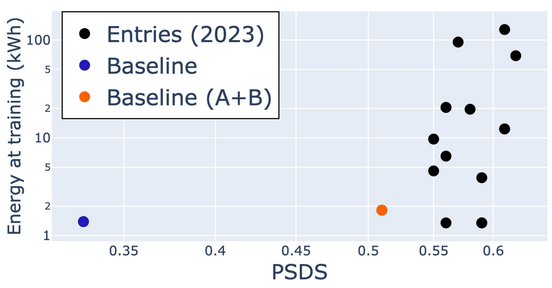
Relation between system complexity, MACs and energy consumption



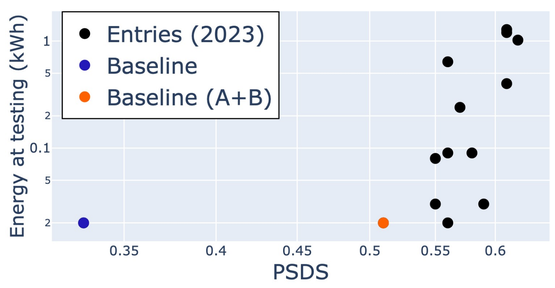
(a) Relation between MACs and energy consumption at training.



(b) System complexity and energy consumption a training.



(c) PSDS and energy consumption at training on 15 top-performing systems.



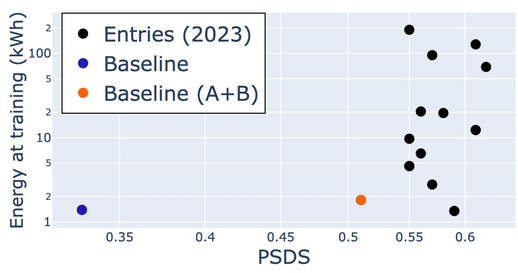
(d) PSDS and energy consumption at testing on 15 top-performing systems.

- The top-performing systems are not the ones consuming the most energy.
- These three different metrics independently are insufficient to provide a comprehensive understanding of the system's footprint.
- Energy at training can be considered as a first (gross) indicator of what would happen at testing phase (but other factors are involved).

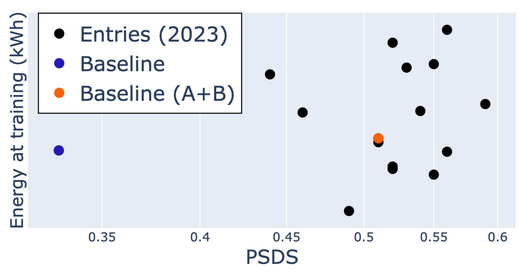
References

[1] Serizel, Romain, Samuele Cornell, and Nicolas Turpault. "Performance Above All? Energy Consumption vs. Performance, a Study on Sound Event Detection with Heterogeneous Data." in ICASSP, IEEE, 2023.

Comparison between ensemble/non-ensemble systems



(a) Relation between PSDS and energy consumption at training for the 15 best ensemble systems.



(b) Relation between PSDS and energy consumption at training for the 15 best not-ensemble systems.

- Ensemble is useful for combining weaker systems for better performance, while a single system can still achieve good results more efficiently.

Thresholding based on energy consumption

	System complexity		MACs		Energy train	
	Max	PSDS_1	Max	PSDS_1	Max	PSDS_1
All	1B	0.59	492 B	0.59	23.00	0.59
Median	6 M	0.59	4 B	0.55	2.33	0.56
25th	5 M	0.55	912 M	0.55	0.99	0.55
All	1B	0.62	21 T	0.62	190.00	0.62
Median	67 M	0.61	72 B	0.60	9.34	0.60
25th	25 M	0.61	8 B	0.58	4.59	0.60

Table: PSDS when system complexity, MACs and training energy are thresholded. The top section is related to not-ensemble systems, the bottom section is related to ensemble systems.

Relation between EW-PSDS and PSDS

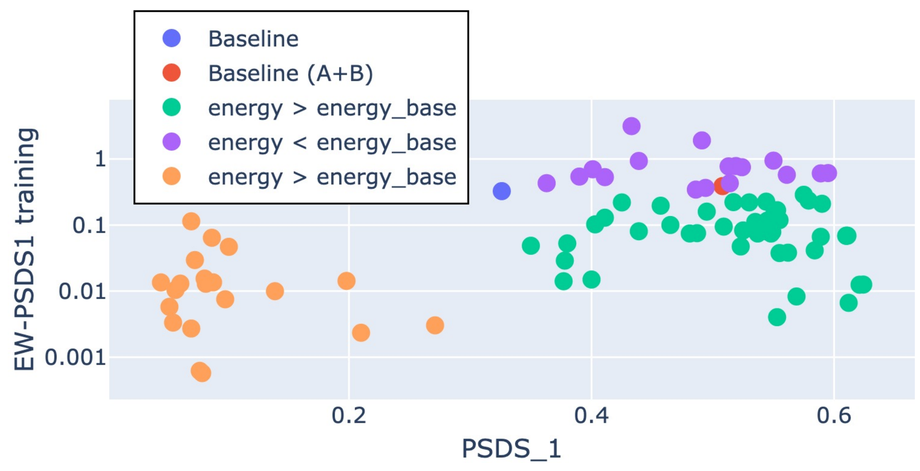


Figure: Relation between PSDS and EW-PSDS at training for 2023 entries.

We should aim to the top-right corner: systems able to right high performance, not underestimating the environmental impact they are going to have.

Conclusions and future works

- Relying on a single metric is insufficient for accurately measuring a system's footprint.
- Systems consuming the most energy (or having the most MACs) do not necessarily outperform less computationally expressive systems.
- There is need for metric(s) capable of taking into account various factors to accurately estimate the energy consumption of deep learning models while taking into account the task-wise performance of the systems.