Data Analytics and Machine Learning Group
Department of Informatics
Technical University of Munich

Accuracy is not the only Metric that matters

Estimating the Energy Consumption of Deep Learning Models

Johannes Getzner, Bertrand Charpentier*, Stephan Günnemann

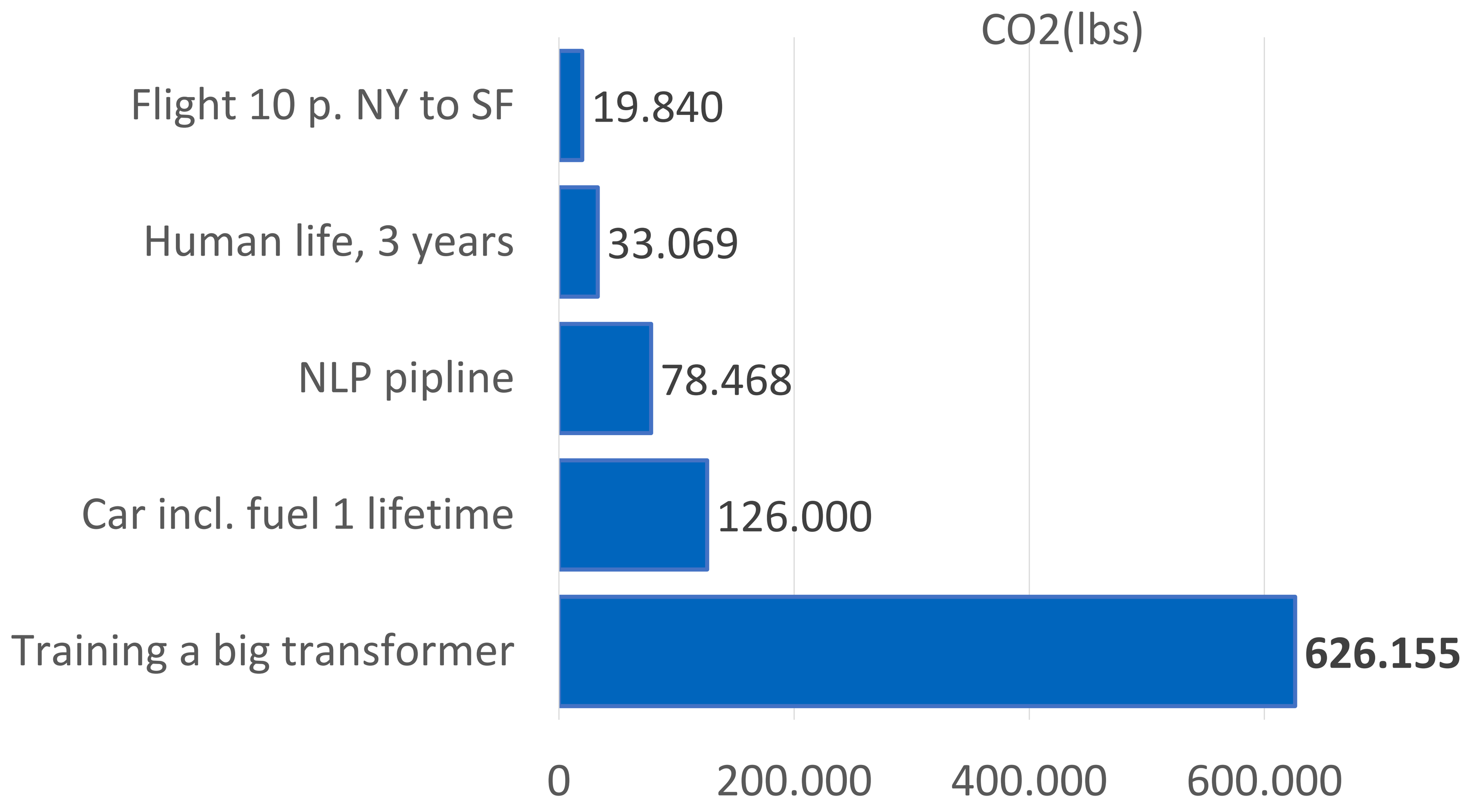
ICLR Climate Change AI

We release…

- ❑ a modular **data-collection process** and a first **dataset on the energy consumption** of various architectures and layer types.
- ❑ a set of energy predictors for different layer types as a simple **energy estimation baseline** for multiple DL architectures.
- ❑ an analysis of the **predictive capabilities of various features sets**, providing key insights into the energy behavior of various DL components.

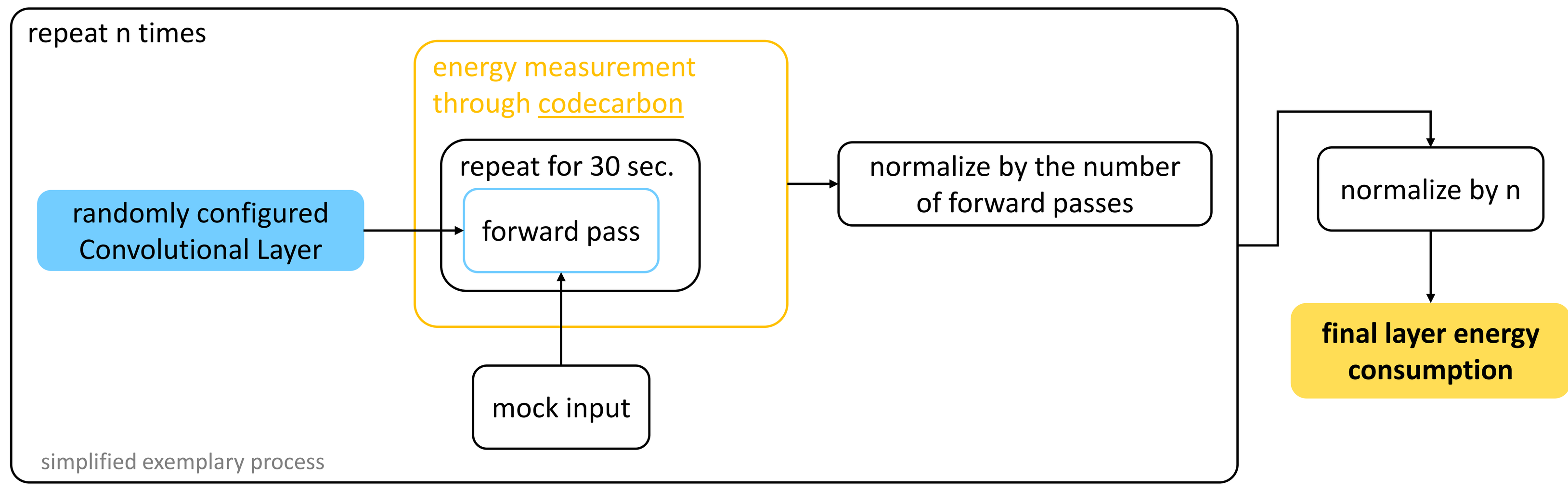
Motivation

Deep Neural Networks consume **astronomical amounts of power**, thus incurring a **large carbon footprint**.



- Training & inference require **power-hungry hardware**, and the electricity often comes from **non-renewable sources**.
- Measures such as **FLOPs may not accurately represent a model's real running costs**.
- Evaluating the cost of a model is generally difficult without running it, especially concerning **environmental impact**.

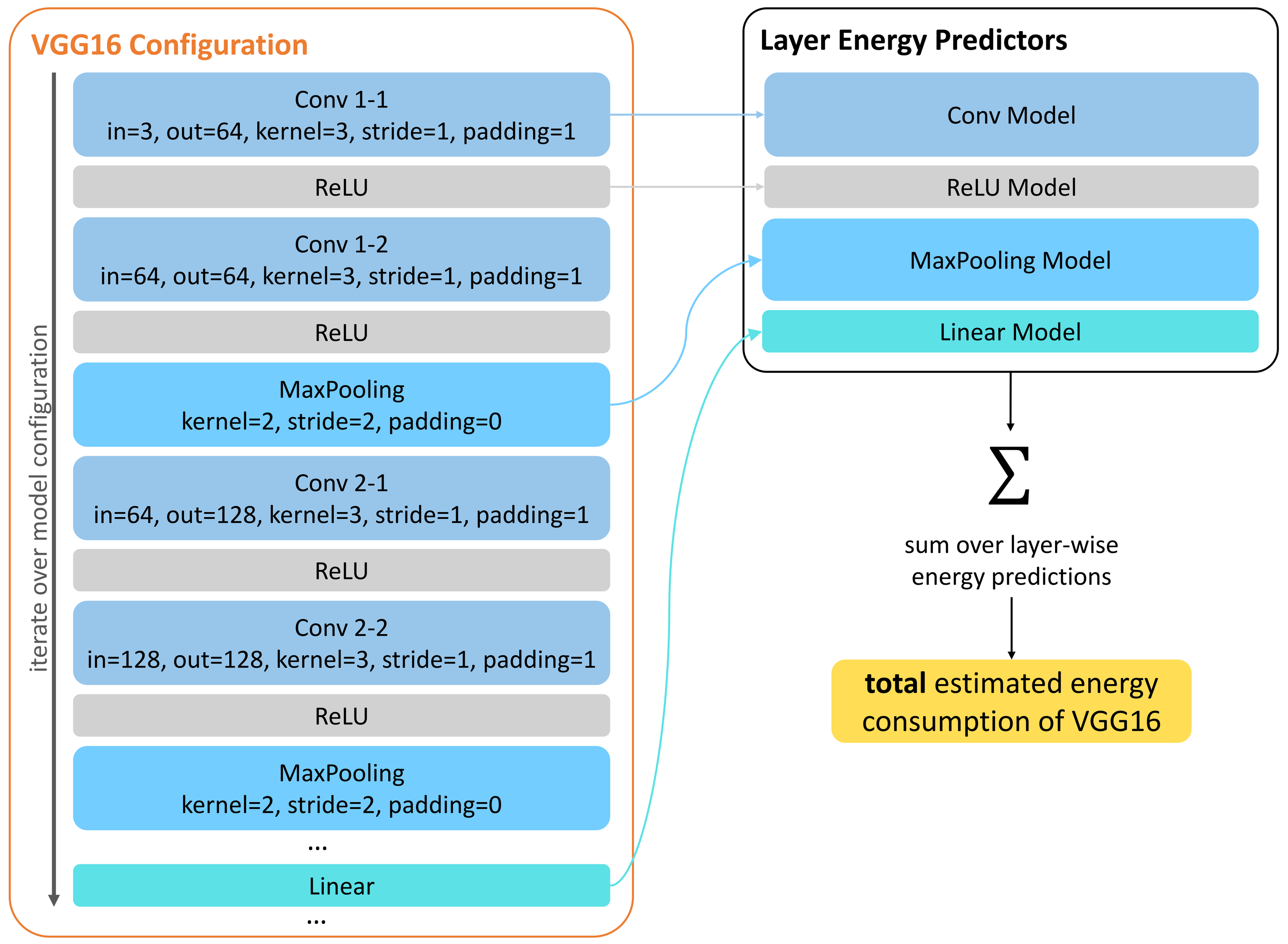
Data Collection Process



- **8 different layer types** have been implemented so far, and we collected approximately **1000 data points** on the energy consumption of each one.
- Due to the modular implementation, more layer types can be easily be added in the future.

Energy Prediction Architecture

We provide energy consumption estimates based only on an architecture's configuration.



- Enables the approximation of the energy consumption **without running** the model.
- Promotes the consideration of **ecological footprint** and the running costs of models, ultimately **raising environmental awareness.**

Energy Prediction Models

Each layer type has a set of **parameters** that can be used to fit the energy estimation model.

Parameters/Features overview

layer	parameters	energy contribution in VGG13
Conv	kernel-size, image-size, in-channels, out-channels, padding, stride	88.42%
MaxPooling	kernel-size, image-size, in-channels, stride	9.14%
Linear	input-size, output-size	1.18%
Activations (ReLU, TanH, Sigmoid, Softmax)	input-size	1.19%

+ batch-size, log-transformed parameters, MAC count

- As no high-order dependencies were found, **polynomial/linear regression models** were chosen to keep the framework lightweight.
- Each model was evaluated concerning its average cross-validation MSE and R^2 score.

Results

The models achieved an excellent performance on the prediction of a random layer's energy consumption.

Layer Energy Predictors

layer	regression model type	chosen features	R^2 score on test set: random layer configs	R^2 score on test set: configs from e.g., VGG 11
Conv	Linear	MAC count	0.9977	0.314
MaxPooling	Polynomial²	all	0.9995	0.559
Linear	Linear	MAC count	0.9992	0.997
ReLU	Polynomial²	MAC count	0.9812	-21.51
TanH	Polynomial²	batch-size, input-size		
Sigmoid	Polynomial²	batch-size, input-size		
Softmax	Polynomial²	batch-size, input-size		

generalization to layers from real architectures is difficult

R^2 score of 0.352 for the total architecture energy predictions of AlexNet and VGG11/13/16.

- Together the models overestimated the total energy consumption slightly.
- The largest contribution to the error comes from the Conv layer.
- More energy-expensive/larger architectures suffer from greater overestimation.

*corresponding author