Measuring Energy Consumption at Method Level for Deep Learning Frameworks

Student: Shang Ziyuan   Supervisor: A/P Ke Yiping, Kelly

Project Objectives

There has been a trend towards increasingly complex deep learning (DL) models to yield better performances. However, as the number of computations required increases, these DL models are becoming more energy-hungry. While there have been some studies addressing the energy consumption of DL models, we believe that energy consumption data of individual methods of popular DL frameworks should be available for DL practitioners to build more energy-efficient models.

Methodology

A client-server architectural setup was devised to measure the energy consumption of individual methods. The source code of a client-side program is first analyzed, and the program is transformed to send POST requests to the server. At the server, the methods under investigation are executed, and energy consumption measurement takes place with the use of Perf and nvidia-smi.
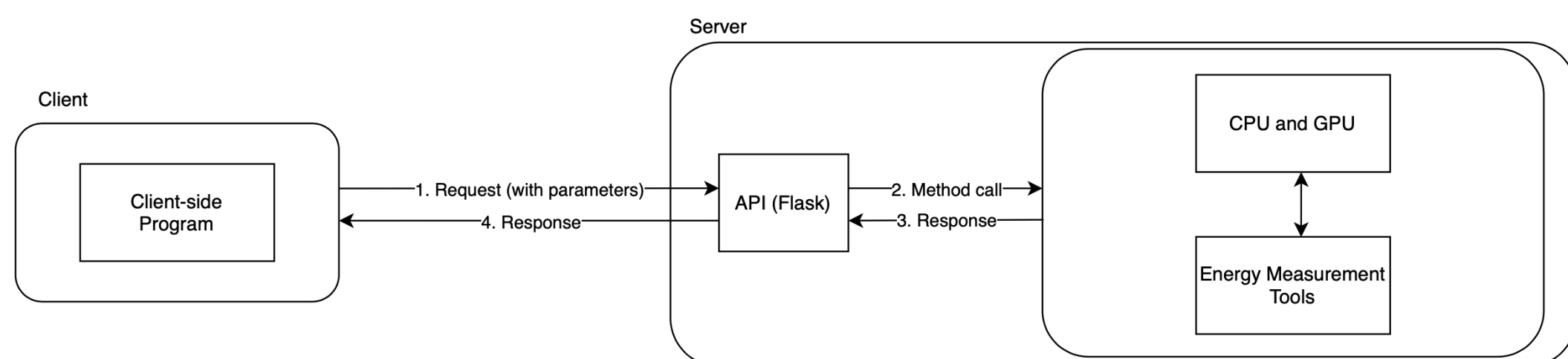
Implementation Results

```
import tensorflow as tf

result = tf.random.uniform([10, 1], minval=1, maxval=1, dtype=tf.float32)
```

```
def custom_method(func, imports, function_to_run, method_object, function_args, function_kwargs, max_wait_secs):
    method_details = {'imports': imports, 'function': function_to_run, 'method_object': method_object,
                      'args': function_args, 'kwargs': function_kwargs, 'max_wait_secs': max_wait_secs}
    data = pickle.dumps(method_details)
    resp = requests.post(url, data=data, headers={'Content-Type': 'application/octet-stream'})
    return func


result = custom_method(
    tf.random.uniform([10, 1], minval=1, maxval=1, dtype=tf.float32), imports='import tensorflow as tf',
    function_to_run='tf.random.uniform(*args, **kwargs)', method_object=None, function_args=['[10, 1]'],
    function_kwargs={'minval': '1', 'maxval': '1', 'dtype': 'tf.float32'}, max_wait_secs=30)
```