BOEING

World Learning
EDUCATION | DEVELOPMENT | EXCHANGE

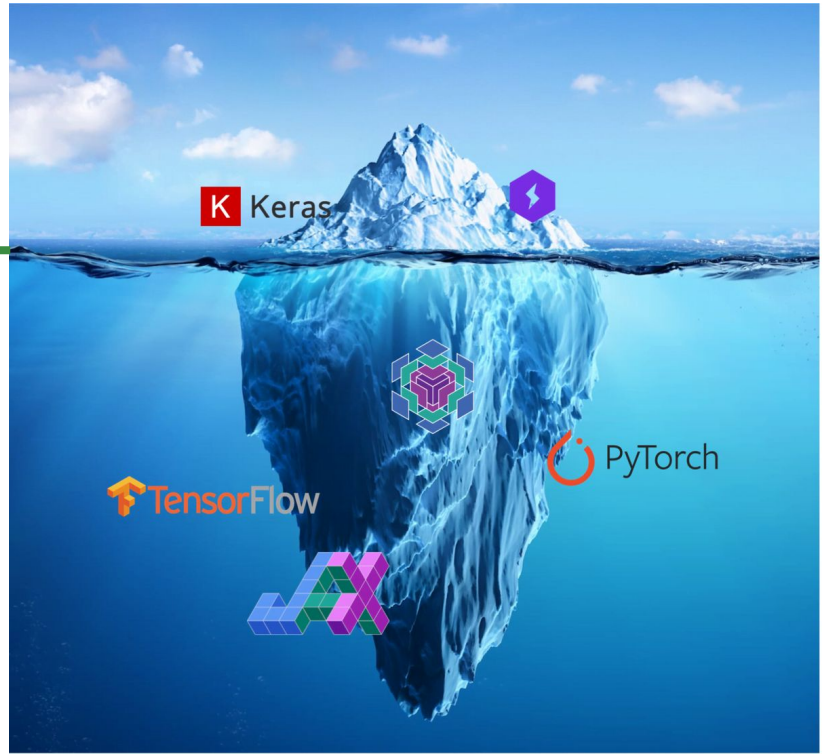
Introduction to

ML Frameworks

Together for a Better Algerian Economy

Overview

High Level

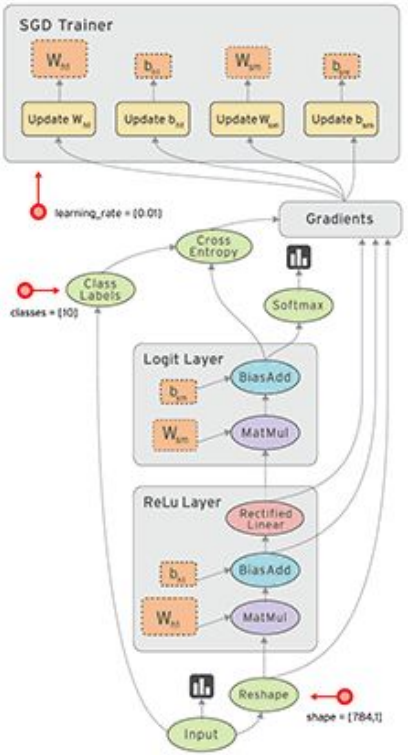


Low Level

TensorFlow + Keras

An open source Deep Learning Library

- Created in 2012
- C++, Python, CUDA
- Develop ML and DL models
- Released by Google Brain in 2015



Google I/O Extended

BOEING

TensorFlow + Keras

An open source Deep Learning Library

- Supports many machine learning tasks (linear regression to complex deep learning architectures)
- Allows for low-level control over the model architecture.
- Higher-level control can be done via Keras, which, with the 2nd version, became a core part of Tensorflow
- Keras 3.0 is again moving away from a Tensorflow-only approach to support a variety of backends: JAX, Tensorflow, and Pytorch.
- Wide selection of platforms available with mobile devices (Tensorflow Lite) or large-scale distributed systems **(Tensorflow Serving)**, production-ready deployment **(TFX - Tensorflow Extended)** or web deployment **(Tensorflow.js for running models in Node.js)**.

Creating Models

Discovering

TF Hub
ML Kit

Data

KYD
TF Data

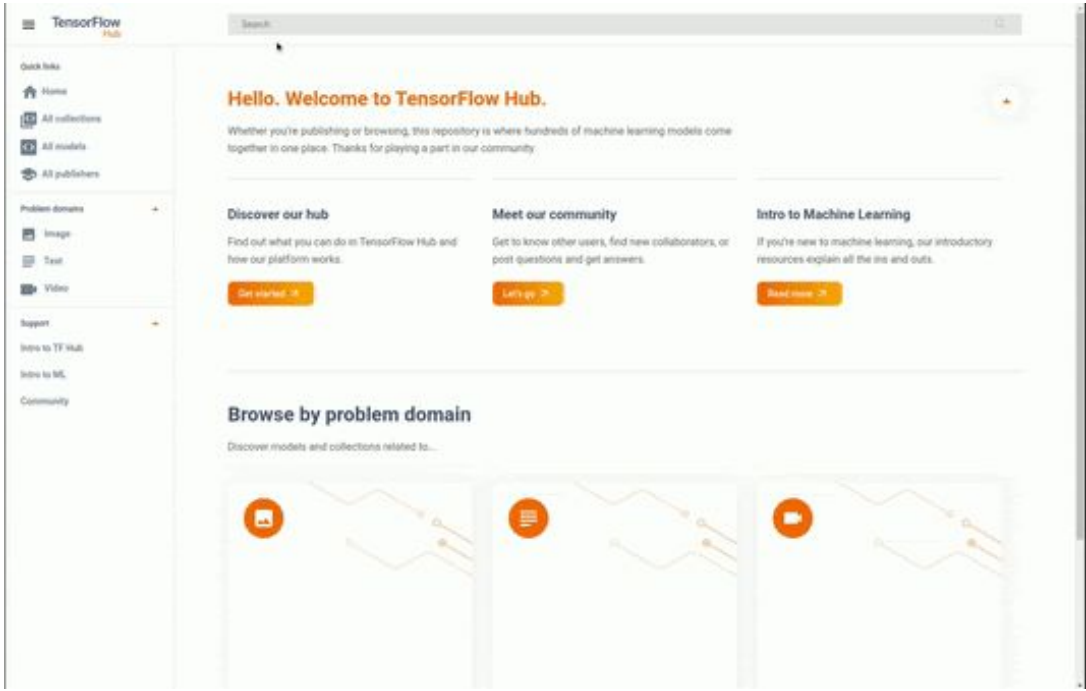
Tooling

TFLMM
TF Cloud
COLAB

TensorFlow + Keras

TF Hub

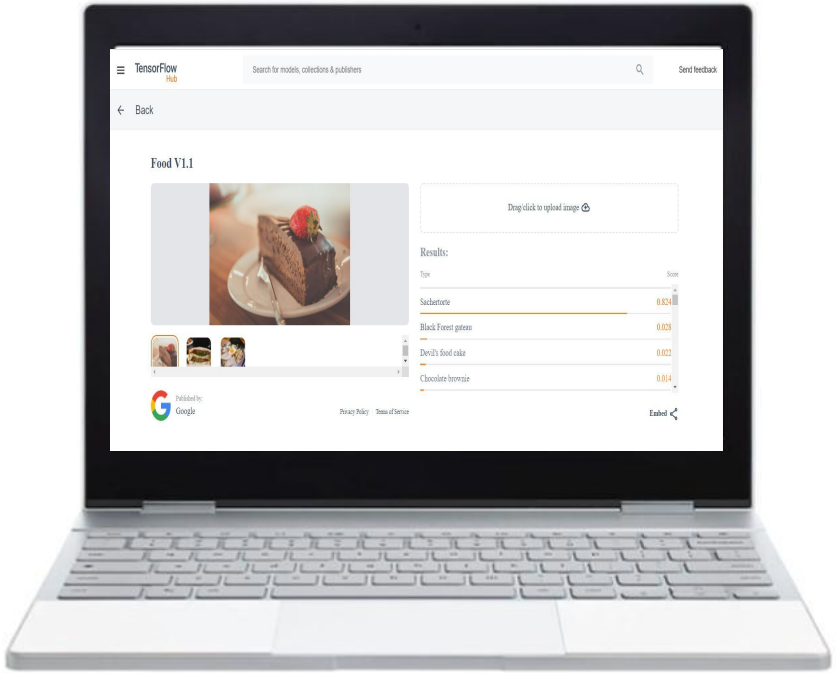
- +10000 pretrained models
- Test models on browser



Google I/O Extended

BOEING

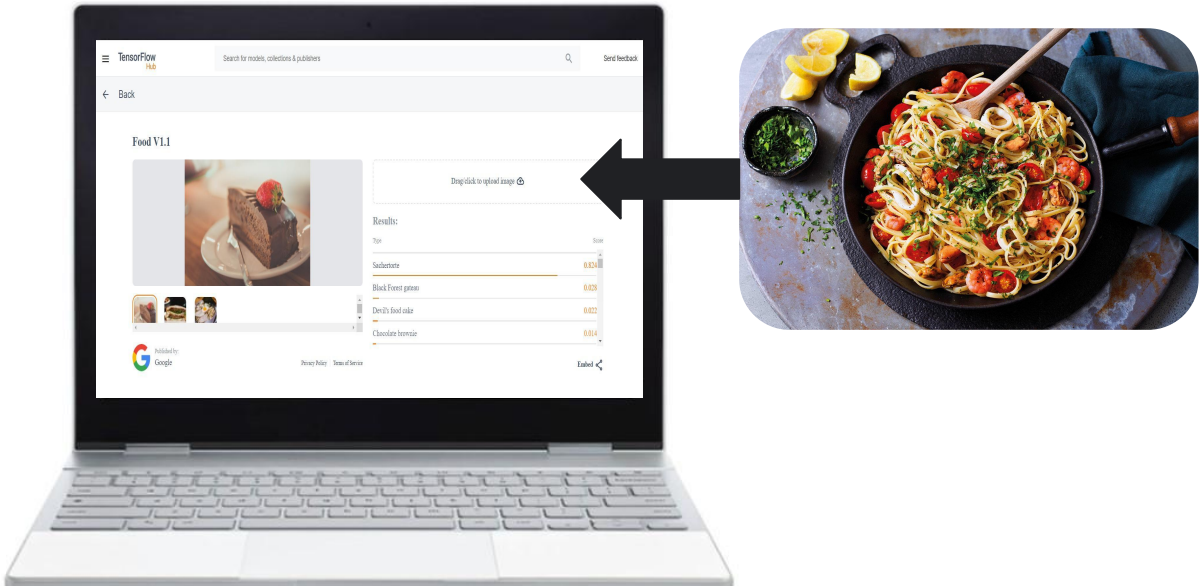
TensorFlow + Keras



Google I/O Extended

BOEING

TensorFlow + Keras



TensorFlow + Keras

Food V1.1




Published by:
Google

Privacy Policy | Terms of Service

Drag/click to upload image

Results:

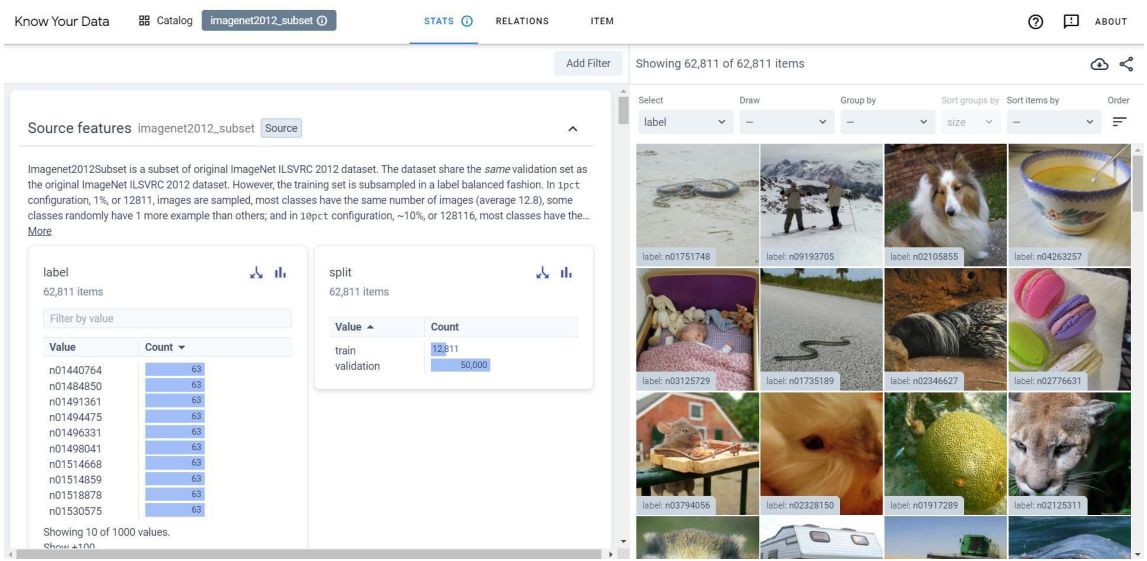
Type	Score
Pad thai	0.203
Carbonara	0.155
Ramen	0.137
Spaghetti	0.123

Embed

Google I/O Extended

BOEING

Know Your Data

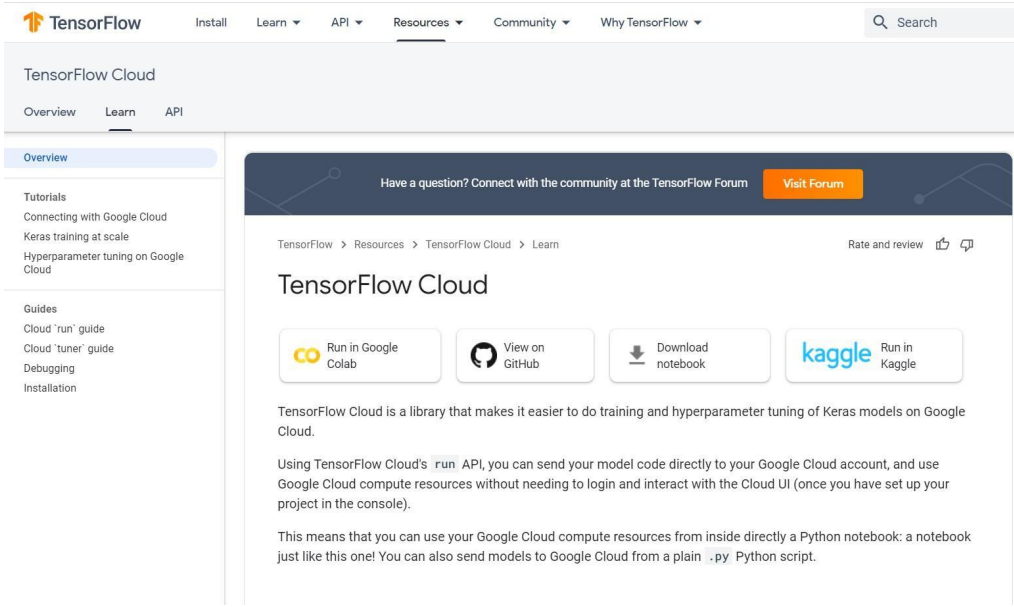


TensorFlow + Keras

TensorFlow Cloud
Easily train with the power of GCP

- Connect your local environment to Google Cloud

Get started:
github.com/tensorflow/cloud



Analysing and Optimizing Models

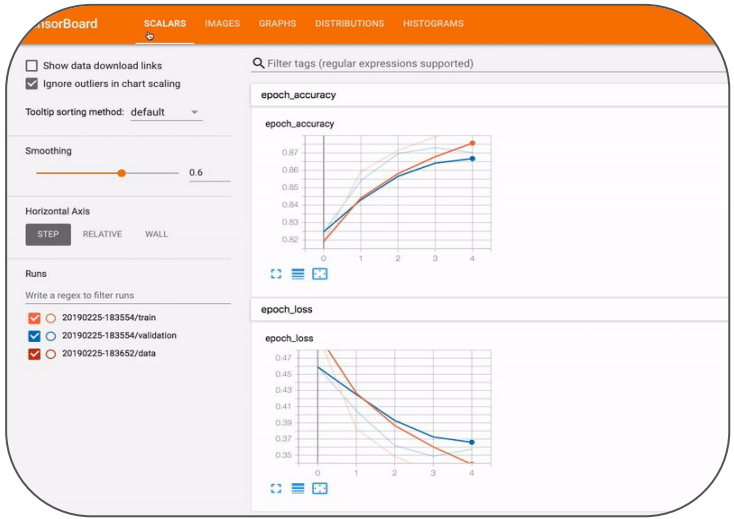
Analysing your Model

TensorBoard
TF Profiler

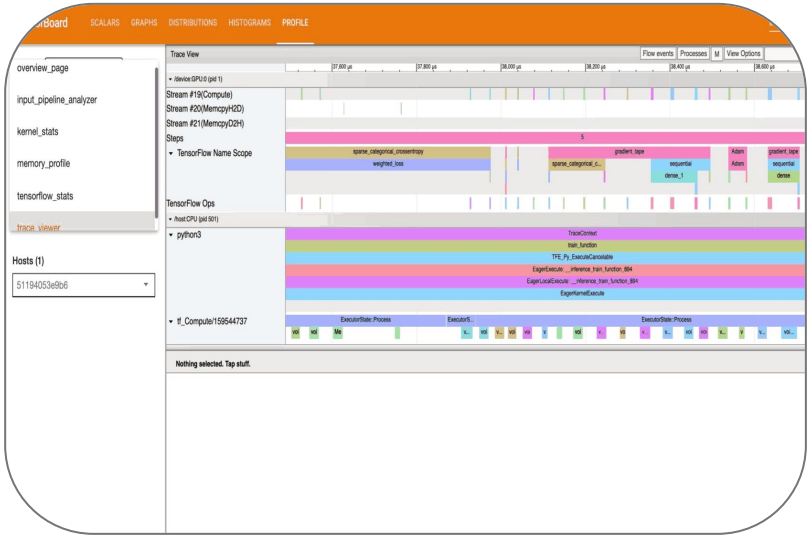
Optimizing your Model

Model Optimization Toolkit
Systrace
Perfetto

TensorFlow + Keras



TensorBoard



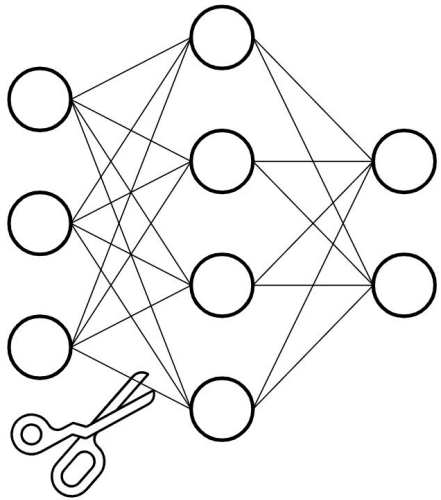
TensorFlow Profiler

Google I/O Extended

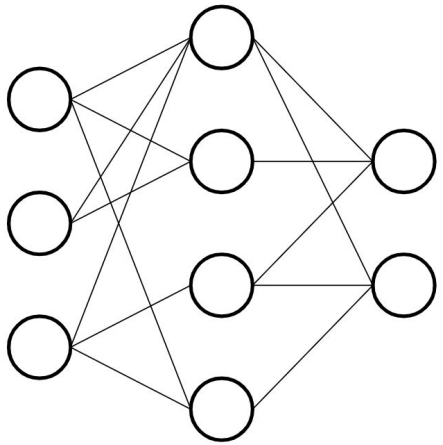
BOEING

Model Optimization Toolkit

- Quantization
- Pruning

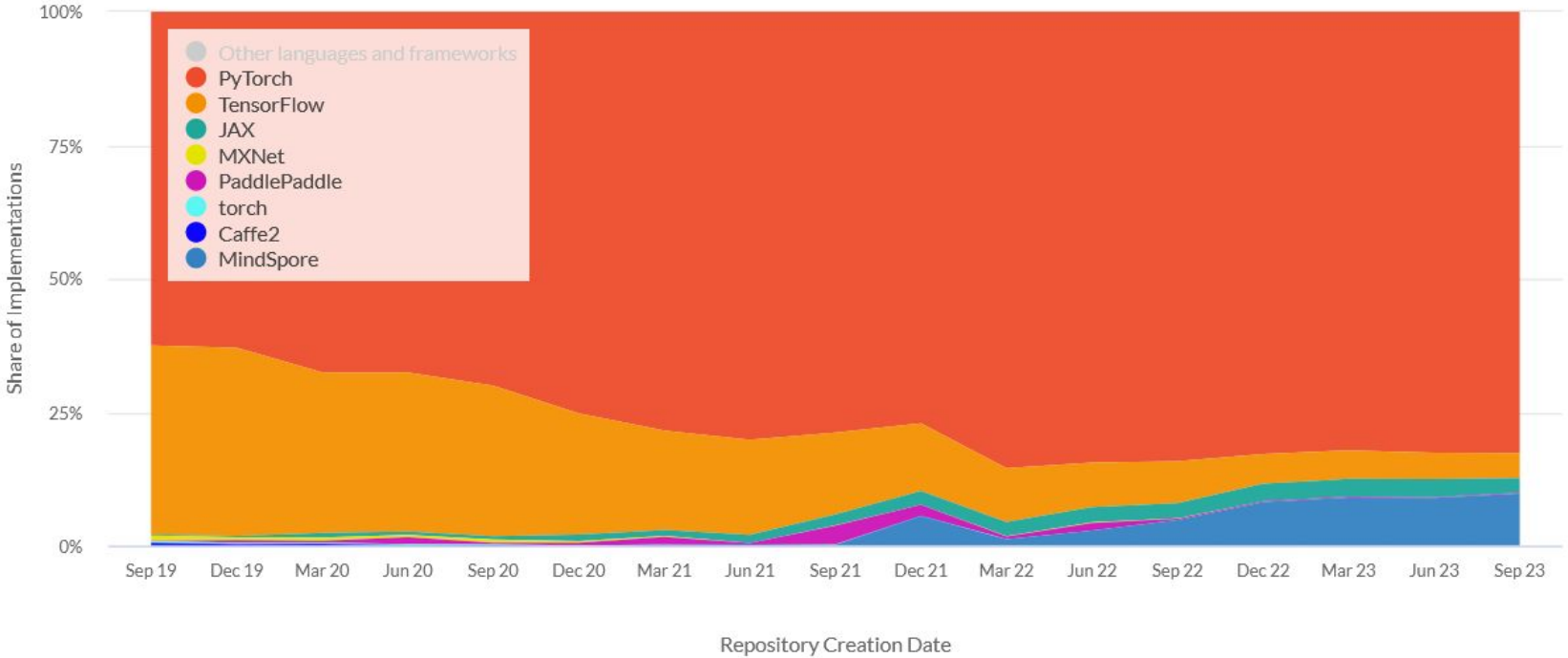


Before pruning



After pruning

PyTorch



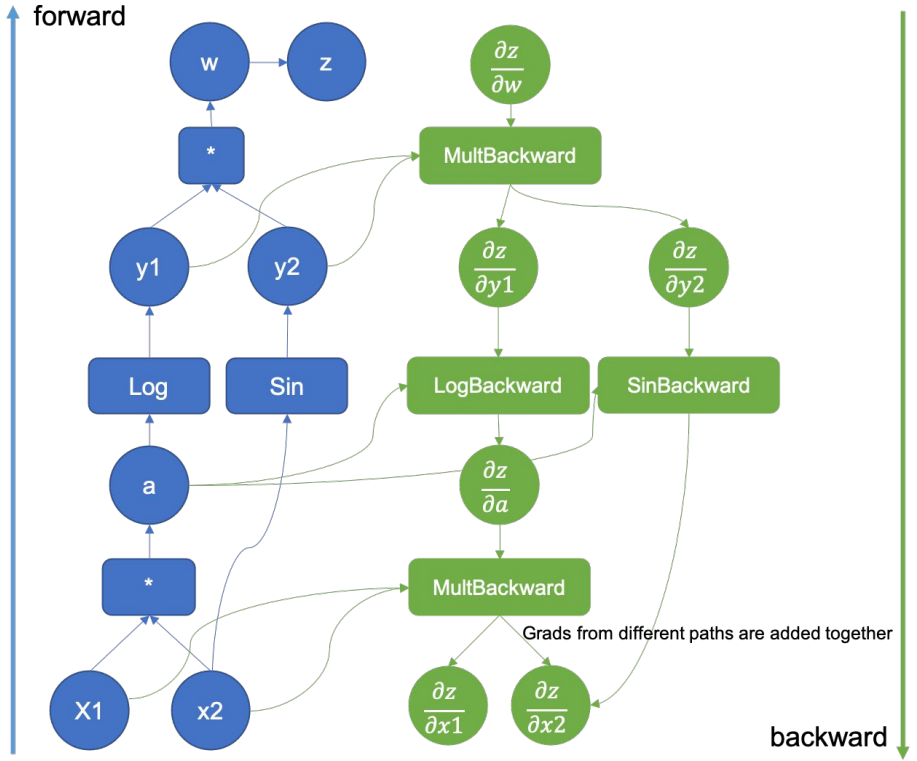
The popularity of frameworks in new Machine Learning papers, source: visio.ai

BOEING

PyTorch

An open source DL Library

- Released in 2016 by Facebook's AI Research Lab (FAIR)
- Dynamic Computation Graph (i.e. Graph built on the fly) and ease of use
- Easier Net debugging and runtime modification
- Auto grad computation
- Growth of specialized tools and libraries (MONAI, Diffusers, Pytorch 3D, fast.ai, etc.)
- Dethroned TF's stable position



BOEING

PyTorch Lightning

An open source DL Library

- High Level Framework built on top of PyTorch
- Abstracts complicated operations that had to be written (i.e., training loops and validation steps) and allows researchers to focus on the core aspects of model development.
- Easy scalability with multiple GPUs and nodes.
- Automatic handling of common tasks like checkpointing, logging or early stopping.
- Even more modular structure, making code easier to read and maintain.

JAX

An open source Numerical Computation Library

- Developed by Google.
- Combines Python flexibility with hardware acceleration.
- Automatic differentiation (autograd) native with Python and NumPy.
- Just-In-Time Compilation (jit), allowing for speedups of Python function into optimised machine code, able to run on CPU, GPU and TPU.
- Extensive vectorisation and parallelism (vmap) enabling efficient code execution across different hardware platforms.

Flax

A High Level Neural Network Library

- Developed by Google.
- Built on top of JAX.
- Provide modular design with ease of customisation.
- Easy manage model parameters, state, and randomness.
- Separate model and training logic, making code easier to debug.

Combining JAX and Flax is the primary weapon of choice by both Deepmind and Google AI.

BOEING

Popularity

More than 75% of new deep learning papers use PyTorch

Differences between Frameworks

Different Initializers of Weights

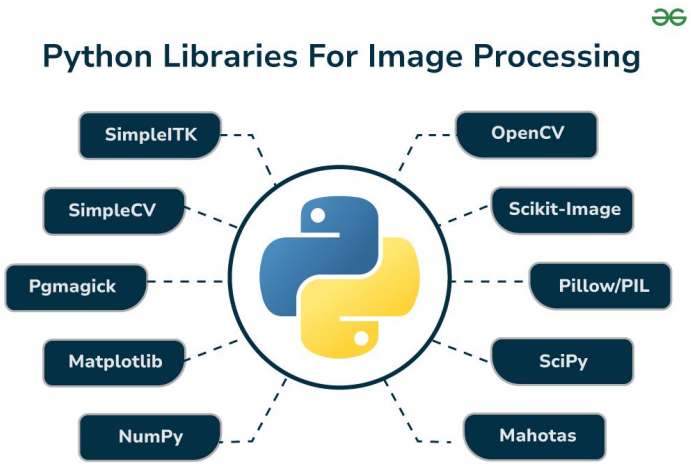
- Tensorflow uses Xavier Uniform initialisation
- Pytorch uses Kaiming Uniform initialisation
- Flax uses LeCun Normal initialisation

Different initializers might greatly influence how the models start the training and in which direction they might go

Differences between Frameworks

Different Processing Backends

- Images and data augmentation handling:
  - Pytorch uses a PIL-based approach
  - Tensorflow/JAX use more native solutions.



BOEING

Differences between Frameworks

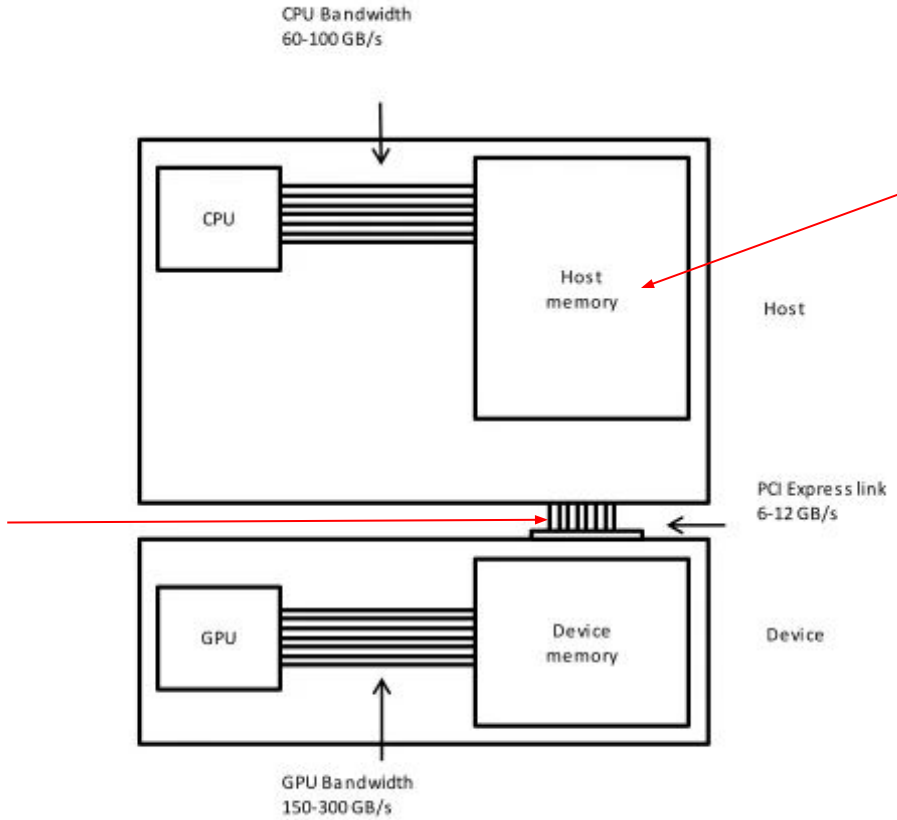
Different GPU Memory Allocation Algorithms

- Pytorch and Pytorch Lightning incrementally allocate memory, allocating more when needed.
- Tensorflow and JAX operate in a greedy fashion, which might cause strange errors when used in the same scope.

Different strategies influence processing time but also allocate more memory when needed, blocking the possibility of using different components that might require GPU.

BOEING

Processing Units

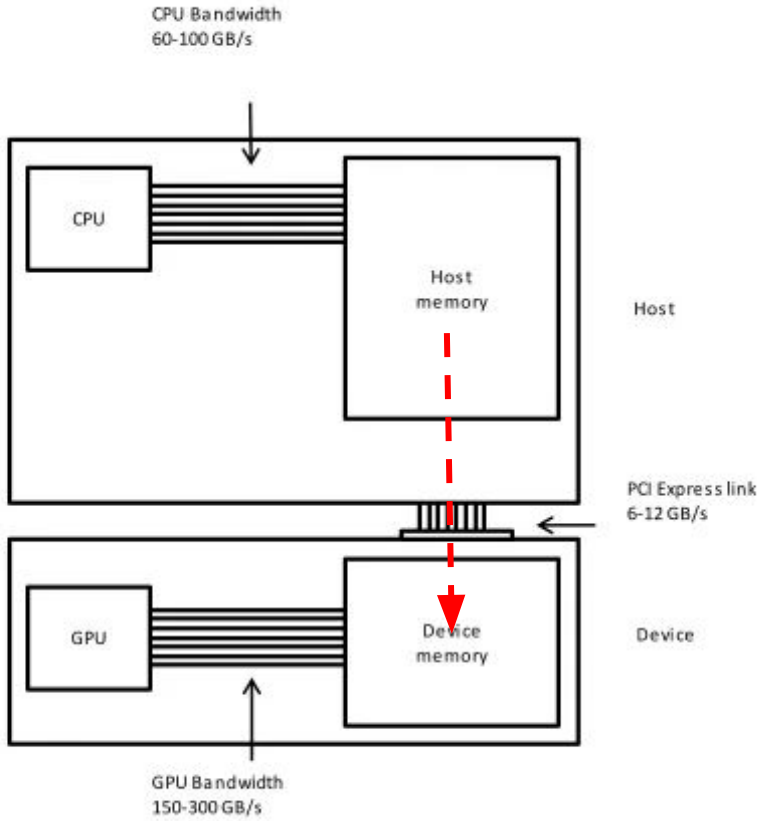


NumPy array only visible to CPU

Physical connection between CPU and GPU but separate memory

BOEING

Processing Units



Data Transfer required from CPU to GPU

BOEING

Conclusion

Which is best ?

- Pytorch is well established.
- Tensorflow has technical debt from previous solutions.
- JAX + Flax still need to show their great potential as their community grows by the day.

BOEING

Which Framework would you like to kickstart with ?

References

What's new in TensorFlow? Google I/O Extended 2021 by Ihssene Brahimi.

https://softwaremill.com/ml-engineer-comparison-of-pytorch-tensorflow-jax-and-flax/

Introduction to GPU programming with cuda and python by Ihssene Brahimi.

World Learning

EDUCATION | DEVELOPMENT | EXCHANGE