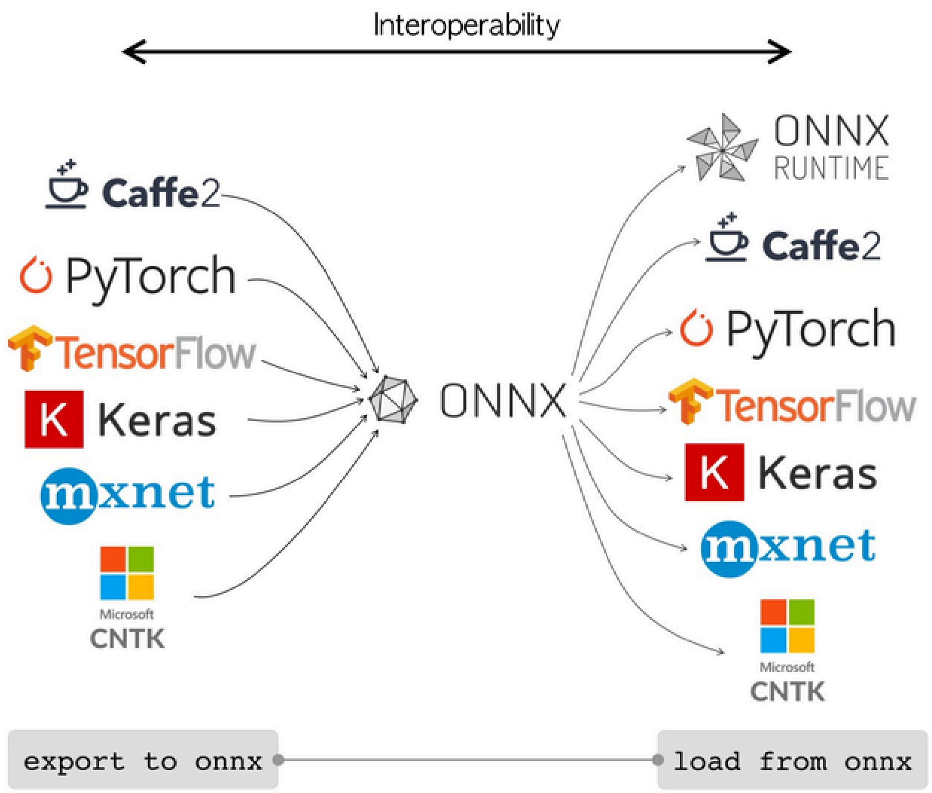

AI Models Preparation for Deployment

Ex 01

3000

To be done individually

Generated by ChatGPT 4

Train a simple neural network using TensorFlow, convert the trained model to different formats (TensorFlow Lite, ONNX, HDF5), and evaluate the size, inference performance, and accuracy of each model.

1. **Dataset Preparation:**

    - Use the CIFAR-10 image dataset, which consists of 60,000 32x32 color images in 10 classes.
    - Load and preprocess the dataset by normalizing the image data and reshaping it for the neural network model.

2. Model Training:

- Build a simple convolutional neural network (CNN) using TensorFlow/Keras.
- Train the model on the CIFAR-10 dataset and save it in .h5 format.

3. Model Conversion:

- Convert the trained model into the following formats:
    - TensorFlow Lite (.tflite)
    - ONNX (.onnx)

- Ensure you install necessary packages for conversions (e.g., tf2onnx, onnx, onnxruntime).

4. Inference Using Different Formats:

- Write separate code snippets to load each model format (.h5, .tflite, .onnx) and perform inference on a sample image from the CIFAR-10 test set.

- Ensure the code outputs the predicted class.

5. Performance Comparison:

- Compare the models in terms of:
  - **Model Size:** Check the file sizes of each format.
  - **Inference Time:** Record the time taken to make predictions for each model format.
  - **Accuracy:** Evaluate the accuracy of each model format on the CIFAR-10 test dataset.

- Discuss which format is more efficient in size, inference speed, and accuracy, and explain why these differences might occur.

6. Report:

- Write a short report summarizing:
    - Model architecture and training process.
    - Challenges faced during model conversion.
    - Comparison results including a table summarizing model size, inference time, and accuracy.
    - Suggestions for improving conversion and performance.

Deliverables:

- Code scripts for each of the above tasks.
- A brief report (1 page) summarizing the findings, with clear tables and graphs where applicable.

Best Wishes