A Scalable Framework for Deep Neural Network Algorithms on Google Cloud Platform

Lingyi Xu*, Questrom School of Business, Boston University

Background

Automatic image classification is an essential topic in texture analysis and cancer diagnosis. Researchers achieve high accuracy (>90%) in binary classification problems using algorithms including *Nearest Neighbor*, *Support Vector Machine*, and *Decision Tree*.

These algorithms have their limitations, especially when it comes to more complex problems such as **multi-class classification**.

One solution is to deploy ***Deep Neural Network* (DNN)** algorithms such as *Convolutional Neural Network*, CNN.

Challenge & Solution

DNN algorithms can automatically extract useful features and significantly improve the accuracy of multi-class problems. However, a DNN model with a complicated structure has many constraints when run on a single machine. We need to find solutions.

Challenges

- Single machines can easily run out of **memory** when training CNN models.
- Single machines require a considerable amount of **time** for model training.
- Complicated models are not well trained on **small datasets**.

Solutions

- Use **GPU** (high throughput) instead of CPU.
- Apply **parallel processing** to increase the computational power horizontally.
- Employ **transfer learning** and **pre-trained models** to eliminate under-fitting or overfitting effects when training small datasets.

Data, Model & Platform

- ➢ **Data**: MNIST dataset, 10 classes
  Kather dataset, 8 classes
- ➢ **Model**: Neural Network models, e.g. VGG16
- ➢ **Platform**: Google Cloud Platform, GCP

Technique

1 Google AI Platform

Google AI Platform provides an integrated tool chain to build and run customized ML applications. It helps scale up model training and prediction in a server-less environment within GCP.

Workflow

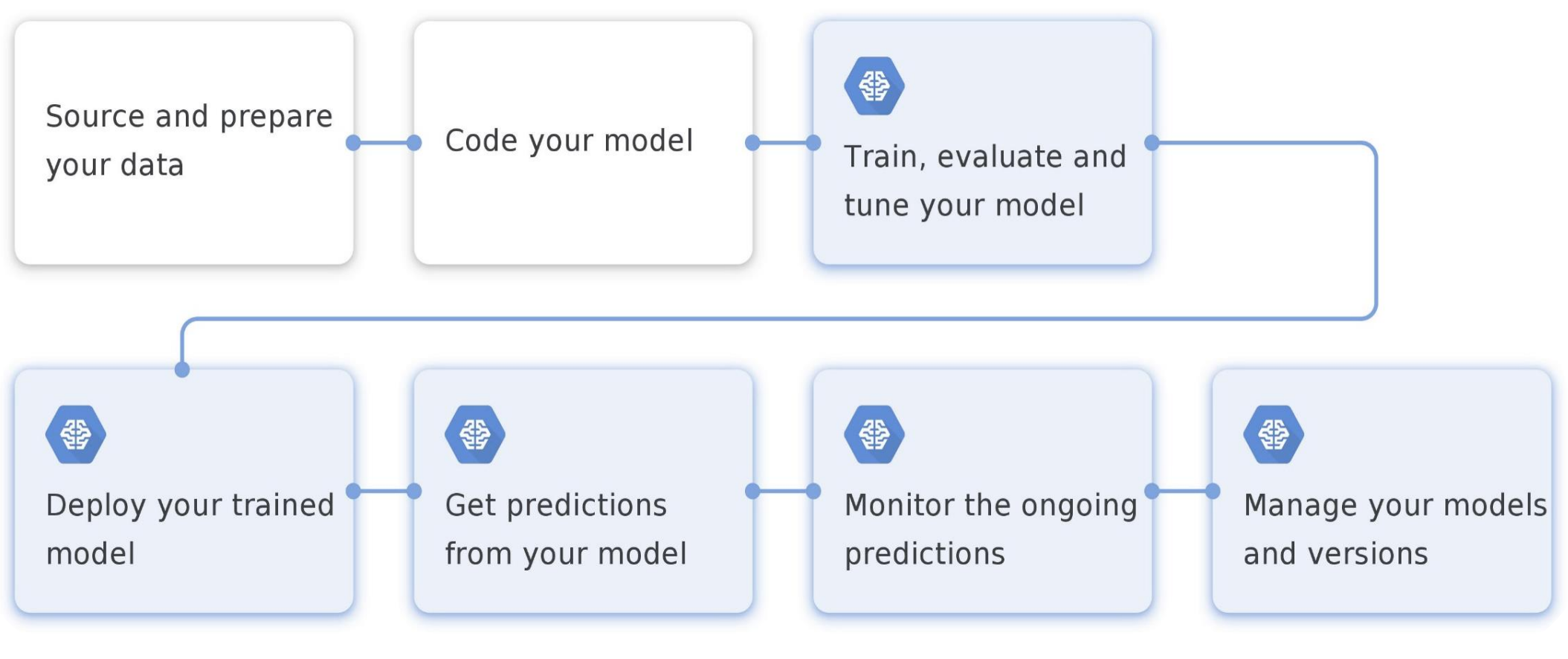

Source and prepare your data
Code your model
Train, evaluate and tune your model
Deploy your trained model
Get predictions from your model
Monitor the ongoing predictions
Manage your models and versions

Figure 1. ML Workflow, with Blue-filled Boxes Managed by AI Platform [1]

Advantage

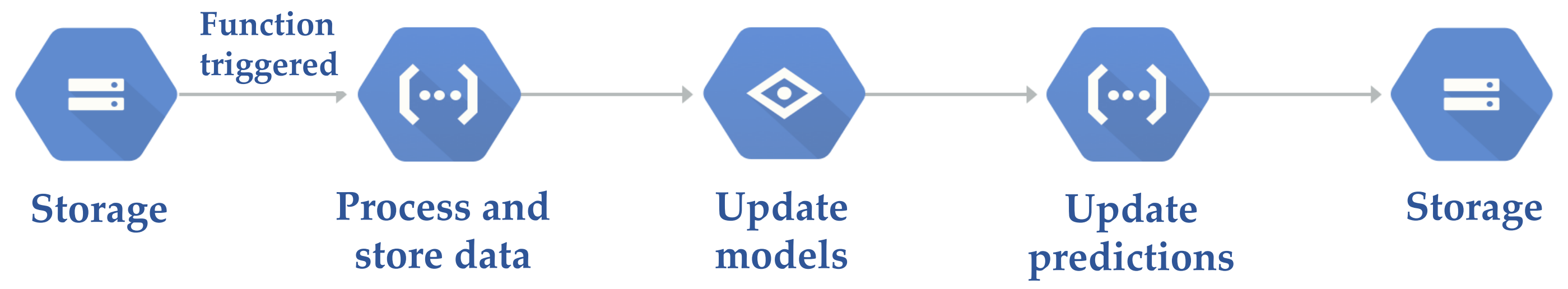
- Customized configuration choices for combination of CPUs and GPUs
- Embedded distributed system and parallel processing mechanism
- Multiple ML frameworks supportive:

  Keras TensorFlow Sikit-learn
- In-line work with Google Cloud Storage and other Google APIs

2 Cloud Dataflow and Data Streaming

Cloud Dataflow is able to transform and enrich data in stream (real time) and batch (historical) modes. **Data streaming** enables models to respond to changes in data, and to meet the need of real-time analysis.

Workflow



Data Streaming can be very useful for processing real-time financial data!

Figure 2. Data Streaming Workflow on GCP [2]

Advantage

- Automated data update detection and resource management
- Models quickly respond to data updates/changes, which enables real-time analysis, e.g. fraud detection
- Integrates data processing techniques with predictive analysis

3 Transfer Learning

Transfer learning is a machine learning method where a model developed for a task (a pre-trained model) is reused for a model on another task. We can fix any number of layers based on our needs.
Pre-trained models can be: CNN models (e.g. VGG16, Inception), general models, any self-trained model

Inception Model

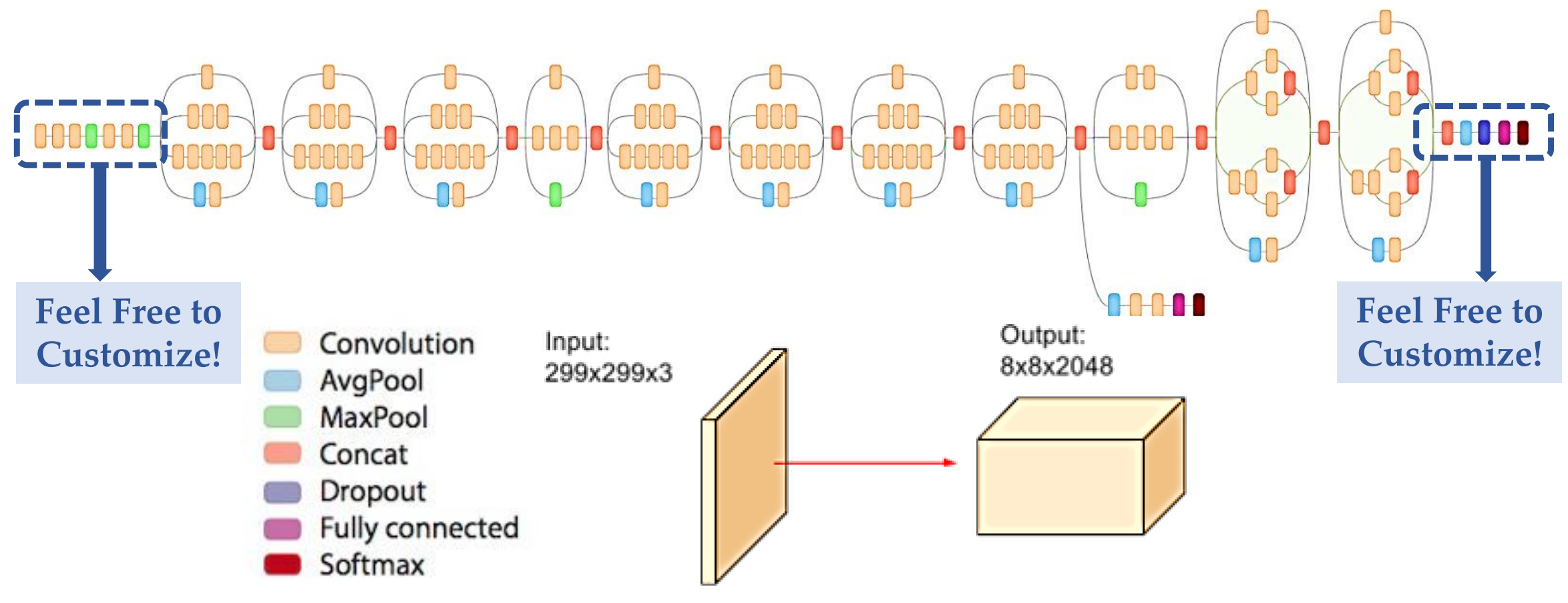


Figure 3. Inception v3 Model [3]

Mechanism

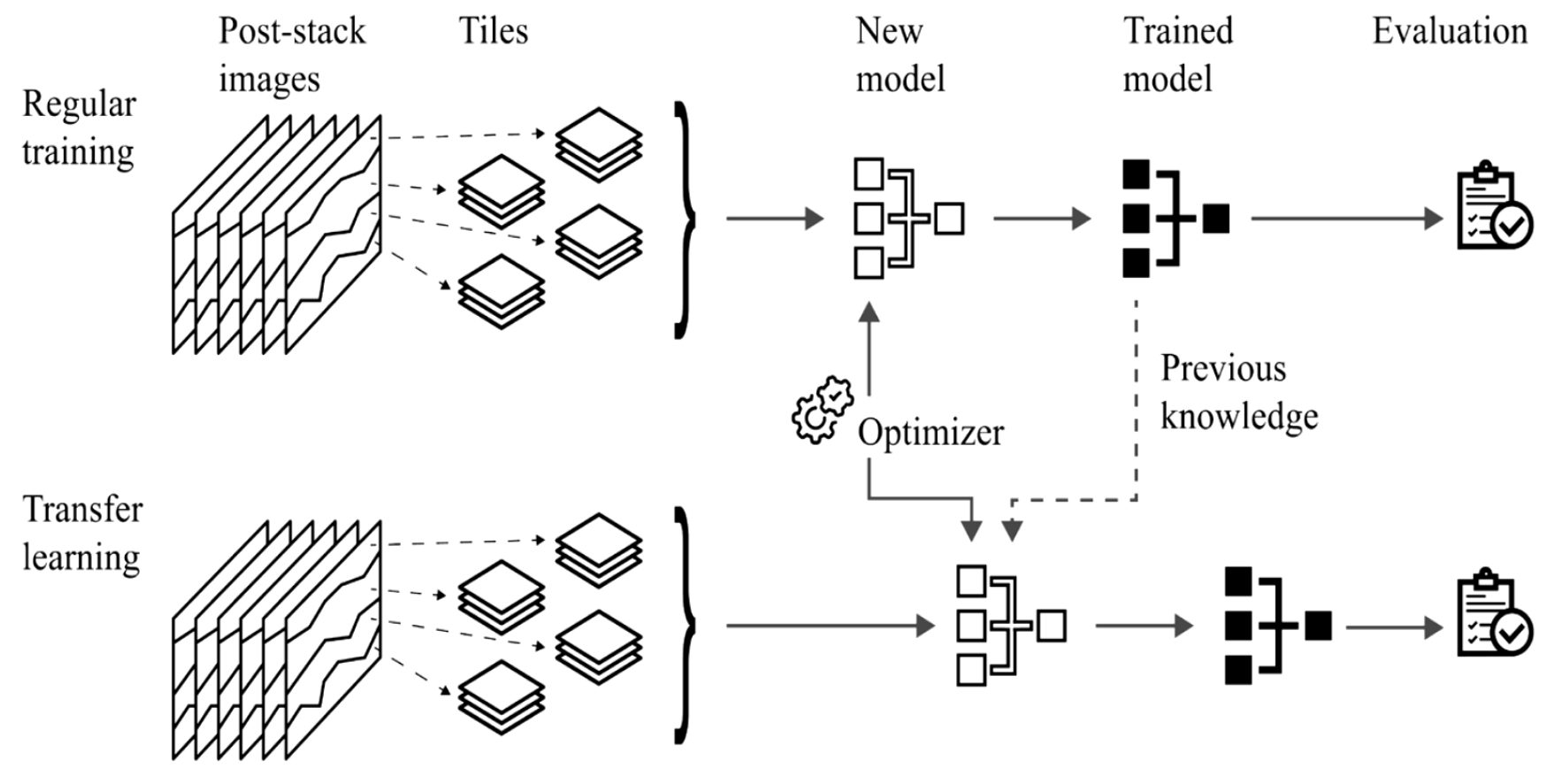


Figure 4. Mechanism of Transfer Learning [4]

Result

➢ **Training time cost** of the MNIST dataset:

Configuration			Training Time
Worker	CPU	GPU(k80)	
1	1	0	-
1	0	1	39 min 42 sec
1	4	0	37 min 15 sec
1	0	4	10 min 48 sec
9	0	1	10 min 20 sec
9	0	4	9 min 4 sec

Table 1. Training Time Comparison with Different Configurations

➢ **Multi-class classification model accuracy** for the Kather dataset (VGG16 as the pre-trained model):

	No CNN, No TL	CNN Without TL	CNN With TL
Test Accuracy	87.40% [5]	88.89%	91.07%

Table 2. Test Accuracy for the Kather Dataset

Conclusion

We have successfully improved model performance and reduced training time cost.

- **Google Cloud Platform** provides a scalable cloud computing framework. This eliminates the constraints of a single machine.
- **Cloud Dataflow and data streaming** speeds up the “Data extraction – Preprocessing – Training – Prediction” workflow.
- **Transfer learning** offers a robust approach to training models on small datasets. It also saves training time without losing too much accuracy.

➢ **Trade-off among time, storage, and accuracy.**
➢ **Choose appropriate techniques based on specific cases and needs.**

Reference

[1] Google AI Platform, Documentation
[2] Google Cloud Functions, Use Cases
[3] Google Cloud TPU, Advanced Guide to Inception v3
[4] D. Chevitarese, D. Szwarcman, R. M. D. Silva, E. V. Brazil. Transfer Learning Applied to Seismic Images Classification. *Search and Discovery Article*, #42285 (2018), October 2018.
[5] J. N. Kather, et al. Multi-class Texture Analysis in Colorectal Cancer Histology. *Scientific Reports*, 6(27988), June 2016.

Contact Information

LinkedIn: linkedin.com/in/olivialingyixu
GitHub: github.com/lingyixu
Email: lingyixu@bu.edu

*This work is under supervision of Professor Mohammad Soltanieh-ha. Great thanks for his guidance and advice on this poster. I also want to thank my teammates in this project, Zhao Huang and Rudrendu Paul.