**CSCIGA.3033 HW 5**

Chenmeinian Guo

The objective of this assignment was to develop and deploy a deep learning training and inference workflow on Google Kubernetes Engine (GKE) using Kubeflow. This report documents my experiences encountered during the implementation and serves as a README along with the codes submitted.

I have two separate Docker containers created: one for training and another for inference. I created these two containers and pushed them to Google Container Registry for deployment on GKE with the following code:

```

# Build Training Image

docker build -t gcr.io/project-id/hw5-training:latest -f Dockerfile.training .

# Build Inference Image

docker build -t gcr.io/project-id/hw5-inference:latest -f Dockerfile.inference .

# Push the images to Google Container Registry

docker push gcr.io/project-id/hw5-training:latest

docker push gcr.io/project-id/hw5-inference:latest

```

Then, I used two YAMLs for training and inference respectively as well.

```

kubectl apply -f training-job.yaml

kubectl apply -f inference-deployment.yaml

```