How to Classify Images with HPCC on Azure

Carina Wang | 2021 LexisNexis Internship

Collect Images

Note: For details on how to collect images (e.g. turning a video into still images, image filetype, cropping images, etc.) refer to my flowchart titled "Collecting Images for Processing with HPCC GNN"

Split images into 2 categories: "train" and "test"

Note: About 90% of the dataset should be in the training folder. 10% should be in the "test" category. The image dataset is now ready.

Choose your Kubernetes environment

Azure → | Local

Note: Reference this link for more information

1. Create an Azure account

3. Install Kubectl

1. Download Docker Desktop → 2. Start Docker + Kubernetes

2. Install Az Cli

4. Create Azure Kubernetes Service (AKS)

3. Create local share such as c:/hpcc-data with the following folders: dalistorage, hpcc-data, queries, sasha, dropzone

Deploy HPCC

Download https://github.com/hpcc-systems/helm-chart.git

cd helm chart

git checkout <version>

Note: example version 8.2.8

Prepare storage

Azure ←

Complete the following steps under this directory: helm/examples/azure

1. Uncomment the following in hpcc/azurefile/values.yaml:
• secretName
• secretNamespace
• all ShareName for each plane

2. cd sa
2.1 edit env-sa
2.2 run "create_sa.sh"
2.3 run "create_secret.sh"

Note:
• 1 - 2.2 only need to be completed once (unless the storage account gets deleted)
• 2.3 - 3.1 need to be completed every time there is a new AKS Kubernetes cluster

Local

1. Under "helm" run: "helm install hpcc-localfile examples/local/hpcc-localfile/"

Note: Right now, <local share> = /c/hpcc-data

3. Go back to examples/azure
3.1 run "helm install azstorage hpcc-azurefile/"

2. Then run: "--set common.host path=<local share>"

Start HPCC

Azure | **Local**

Under helm directory:
run "helm install hpcc hpcc/ --set global.image.version=8.2.8 --set global.image.name=platform-gnn -f examples/azure/values-retained-azurefile.yaml"

Under helm directory:
run "helm install hpcc hpcc/ --set global.image.version=8.2.8 --set global.image.name=platform-gnn -f examples/local/values-localfile.yaml"

Note: Verify that HPCC is ready with "kubectl get pods" and "kubectl get service"

Prepare GNN Development

ECL IDE | **VS Code**

Download IDE

Install HPCC Clienttools

Install ML and GNN bundles

Install VS Code ECL Language Extension.

Wait for bundles to be listed. Then, click on the ones you want to download.

Note: For small datasets, use ECL Watch to load and spray data. For large datasets (1,000+ images) use the "kubectl cp" command to load the images, and use the "dfu plus" command to spray the data.

Download GNN Tutorial code from my GitHub project

Configure launch.json and settings.json in VS Code

Load and spray image data onto HPCC cluster

Define datatype based on sprayed data

Develop GNN model and train with data

Predict with a new image