Cloud AutoML Vision Object Detection

Training Edge exportable models

You create a custom model by training it using a prepared dataset (https://cloud.google.com/vision/automl/docs/create-datasets). AutoML API uses the items from the dataset to train the model, test it, and evaluate (https://cloud.google.com/vision/automl/docs/evaluate) its performance. You review the results, adjust the training dataset as needed, and train a new model using the improved dataset.

Training a model can take several hours to complete. The AutoML API enables you to check the status (https://cloud.google.com/automl/docs/reference/rest/v1/projects.locations.operations/get) of training.

Since AutoML Vision creates a new model each time you start training, your project may include numerous models. You can get a list of the models in your project (https://cloud.google.com/vision/automl/docs/models#list-models) can delete models (https://cloud.google.com/vision/automl/docs/models#delete-model) you no longer need. Alternatively, you can use the Cloud AutoML Vision UI to list and delete models created via the AutoML API that you do not need anymore.

Note:

- Unless otherwise specified in applicable terms of service or documentation, custom models created in Cloud AutoML products cannot be exported.
- The maximum lifespan for a custom model is 18 months as of the GA release. You must create and train a new model to continue classifying content after that amount of time.
- Edge models are optimized for inference on an Edge device. Consequently, Edge model accuracy ***will differ*** from Cloud model accuracy.

Models are based on state-of-the-art research (https://ai.googleblog.com/2018/08/mnasnet-towards-automating-design-of.html) at Google. Your model will be available as a TF Lite package. For more information about how to integrate a TensorFlow Lite model using the TensorFlow Lite SDK reference the following links for iOS (https://www.tensorflow.org/lite/demo_ios) and Android (https://www.tensorflow.org/lite/demo_android).

Training Edge models

When you have a dataset with a solid set of labeled training items, you are ready to create and train your custom Edge model.

TensorFlow serving and TF Lite models

When training Edge models you can specify three distinct values in the `modelType` (https://cloud.google.com/automl/docs/reference/rest/v1/projects.locations.models#ImageObjectDetectionModelMetadata) field depending on your model needs:

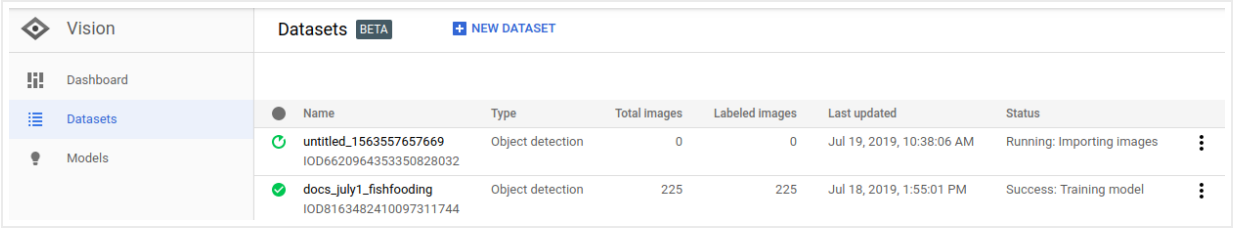
- `mobile-low-latency-1` for low latency,
- `mobile-versatile-1` for general purpose usage, or
- `mobile-high-accuracy-1` for higher prediction quality.

The model type will also be shown in the API request response.

WEB UI | REST & CMD LINE | MORE

1. Open the Cloud AutoML Vision Object Detection UI (https://console.cloud.google.com/vision).

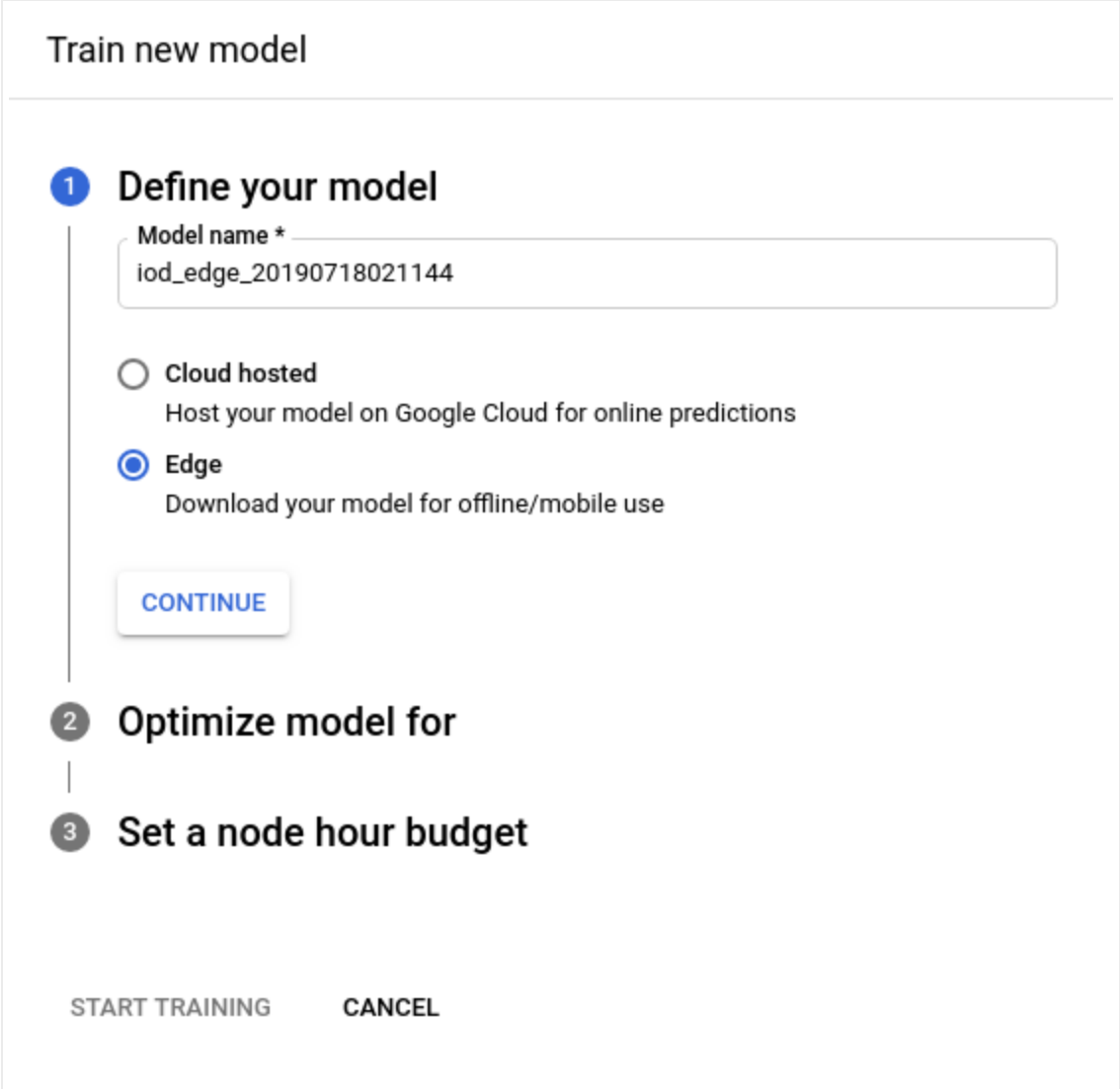
   The **Datasets** page shows the available datasets for the current project.



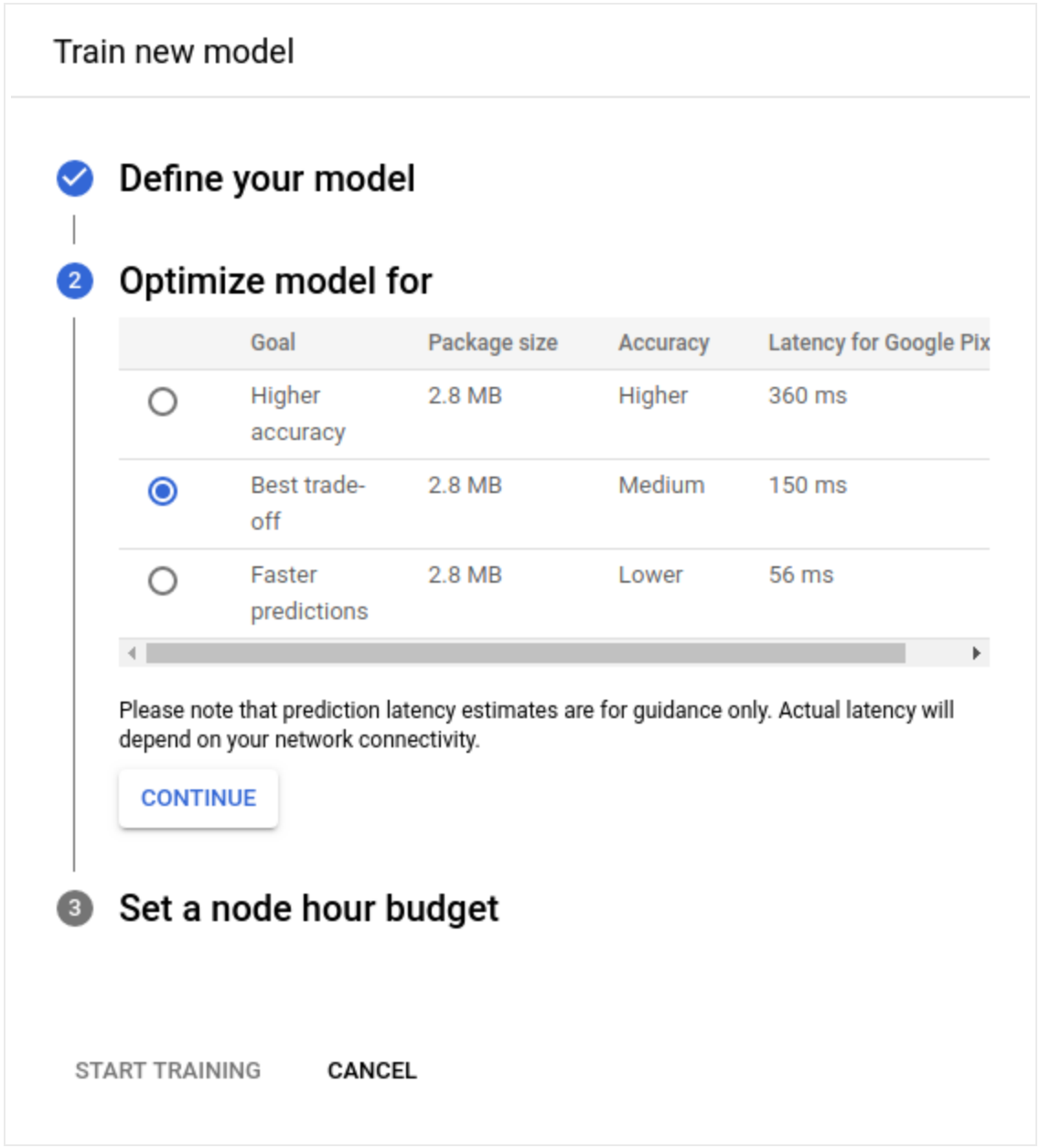
2. Select the dataset you want to use to train the custom model.
3. When the dataset is ready, select the **Train** tab and **Train new model** button.

   This will open a **"Train new model"** side window with training options.
4. From the training **Define your model** section, change the model name (or use the default value) and select ◉ **Edge** as the model type. After selecting to train an Edge

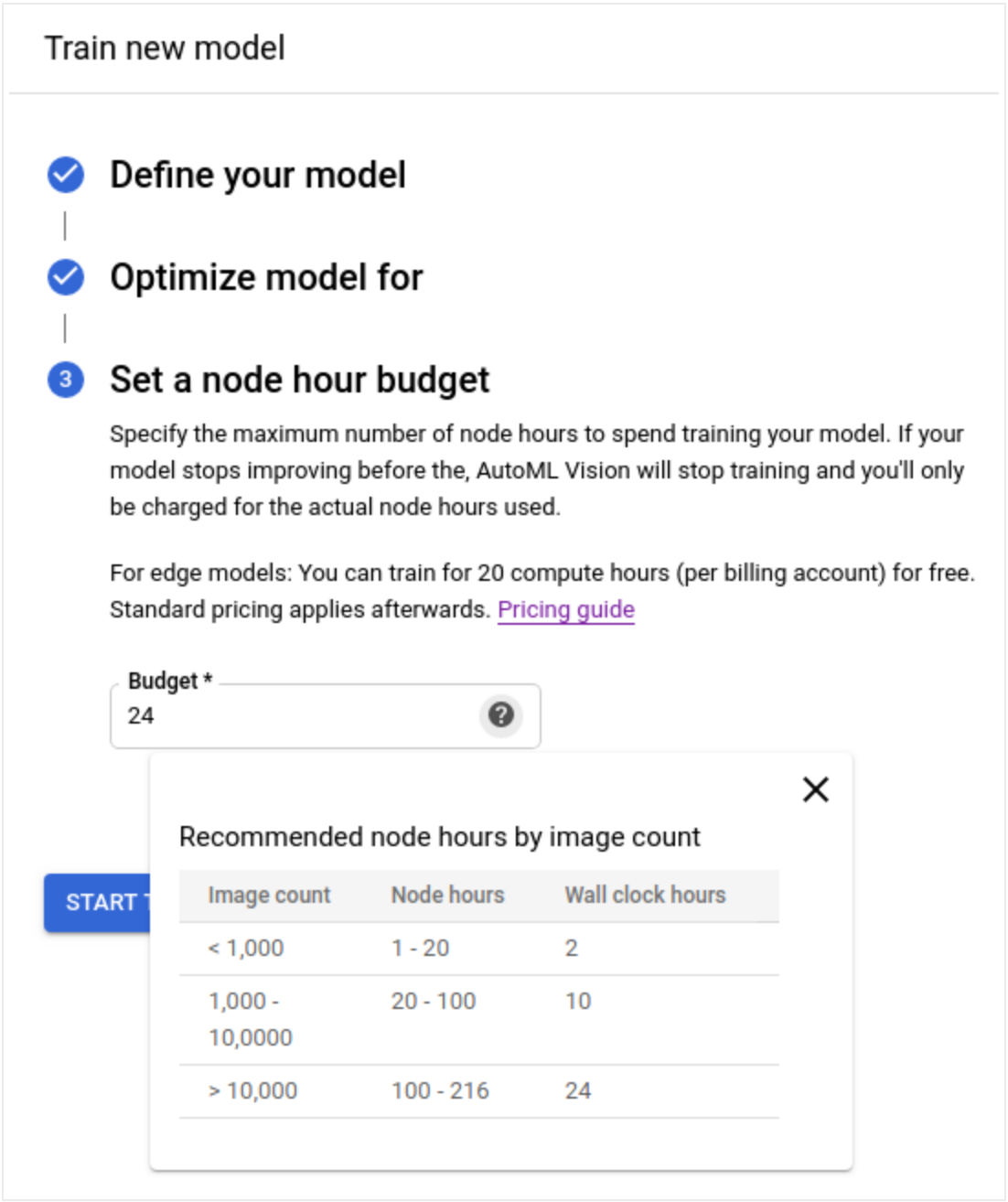
model select **Continue**.



5. In the following **Optimize model for** section, select your desired optimization criterion: ◉ **Higher accuracy**, **Best tradeoff**, or **Faster prediction**. After selecting the optimization specification select **Continue**.



6. In the following **Set a node hour budget** section use the recommended node hour budget, or specify a different value.



Node hour budget: Recommended training time is calculated based on:

- model learning curves
- training dataset size

If model training converges *before* recommended or custom time selected the system allows for early stopping; this means you are *only charged for the time it takes to train the model*. **You are encouraged to use the recommended amount of hours to avoid using more resources than necessary and reduce billing costs**.

7. Select **Start training** to begin model training.

Training a model can take several hours to complete. After the model is successfully trained, you will receive a message at the e-mail address that you used for your Google Cloud Platform project.

List operations status

You can list your project's operations, and filter results.

REST & CMD LINE | C# | GO | MORE

Before using any of the request data below, make the following replacements:

- ***project-id***: your GCP project ID.

HTTP method and URL:

```
GET https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/ope
```

To send your request, choose one of these options:

CURL | POWERSHELL

Note: Ensure you have set the **GOOGLE_APPLICATION_CREDENTIALS** (https://cloud.google.com/docs/authentication/production) environment variable to your service account private key file path.

Execute the following command:

```
curl -X GET \
-H "Authorization: Bearer "$(gcloud auth application-default print-access-token)
https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/opera
```

The output you see will vary depending on the operations you have requested.

You can also filter the operations returned by using select query parameters (`operationId`, `done`, and `worksOn`). For example, to return a list of operations that have finished running modify the URL:

```
GET https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/oper
```

Getting the status of an operation

REST & CMD LINE | C# | GO | MORE

Before using any of the request data below, make the following replacements:

- ***project-id***: your GCP project ID.
- ***operation-id***: the ID of your operation. The ID is the last element of the name of your operation. For example:
  - operation name: `projects/project-id/locations/location-id/operations/IOD5281059901324392598`
  - operation id: `IOD5281059901324392598`

HTTP method and URL:

```
GET https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/ope
```

To send your request, choose one of these options:

CURL | POWERSHELL

Note: Ensure you have set the GOOGLE_APPLICATION_CREDENTIALS (https://cloud.google.com/docs/authentication/production) environment variable to your service account private key file path.

Execute the following command:

```
curl -X GET \
-H "Authorization: Bearer "$(gcloud auth application-default print-access-token)
https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/opera
```

You should see output similar to the following for a completed **import operation**:

```
{
  "name": "projects/project-id/locations/us-central1/operations/operation-id",
  "metadata": {
    "@type": "type.googleapis.com/google.cloud.automl.v1.OperationMetadata",
    "createTime": "2018-10-29T15:56:29.176485Z",
    "updateTime": "2018-10-29T16:10:41.326614Z",
```

```
    "importDataDetails": {}
  },
  "done": true,
  "response": {
    "@type": "type.googleapis.com/google.protobuf.Empty"
  }
}
```

You should see output similar to the following for a completed **create model operation**:

```
{
  "name": "projects/project-id/locations/us-central1/operations/operation-id",
  "metadata": {
    "@type": "type.googleapis.com/google.cloud.automl.v1.OperationMetadata",
    "createTime": "2019-07-22T18:35:06.881193Z",
    "updateTime": "2019-07-22T19:58:44.972235Z",
    "createModelDetails": {}
  },
  "done": true,
  "response": {
    "@type": "type.googleapis.com/google.cloud.automl.v1.Model",
    "name": "projects/project-id/locations/us-central1/models/model-id"
  }
}
```

Cancelling an Operation

You can cancel an import or training task using the operation ID.

REST & CMD LINE

Before using any of the request data below, make the following replacements:

- ***project-id***: your GCP project ID.
- ***operation-id***: the ID of your operation. The ID is the last element of the name of your operation. For example:
  - operation name: `projects/project-id/locations/location-id/operations/IOD5281059901324392598`
  - operation id: **`IOD5281059901324392598`**

HTTP method and URL:

```
POST https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/op
```

To send your request, choose one of these options:

CURL POWERSHELL

Note: Ensure you have set the **GOOGLE_APPLICATION_CREDENTIALS** (https://cloud.google.com/docs/authentication/production) environment variable to your service account private key file path.

Execute the following command:

```
curl -X POST \
-H "Authorization: Bearer "$(gcloud auth application-default print-access-token)
-H "Content-Type: application/json; charset=utf-8" \
-d "" \
https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/opera
```

You will see an empty JSON object returned from a successful request:

```
{}
```

Getting information about a model

When training is complete, you can get information about the newly created model.

The examples in this section return the basic metadata about a model. To get details about a model's accuracy and readiness, see the "Evaluating models" topic.

REST & CMD LINE C# GO MORE

Before using any of the request data below, make the following replacements:

- ***project-id***: your GCP project ID.

- ***model-id***: the ID of your model, from the response when you created the model. The ID is the last element of the name of your model. For example:
  - model name: `projects/`***`project-id`***`/locations/`***`location-id`***`/models/IOD4412217016962778756`
  - model id: `IOD4412217016962778756`

HTTP method and URL:

```
GET https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/mod
```

To send your request, choose one of these options:

CURL POWERSHELL

Note: Ensure you have set the **GOOGLE_APPLICATION_CREDENTIALS** (https://cloud.google.com/docs/authentication/production) environment variable to your service account private key file path.

Execute the following command:

```
curl -X GET \
-H "Authorization: Bearer "$(gcloud auth application-default print-access-token)
https://automl.googleapis.com/v1/projects/project-id/locations/us-central1/model
```

You should receive a JSON response similar to the following:

```
{
  "name": "projects/project-id/locations/us-central1/models/model-id",
  "displayName": "display-name",
  "datasetId": "dataset-id",
  "createTime": "2019-07-29T17:16:34.476787Z",
  "deploymentState": "UNDEPLOYED",
  "updateTime": "2019-07-29T18:30:13.601461Z",
  "imageObjectDetectionModelMetadata": {
    "modelType": "mobile-low-latency-1",
    "nodeQps": -1,
    "stopReason": "MODEL_CONVERGED",
    "trainBudgetMilliNodeHours": "24000",
    "trainCostMilliNodeHours": "861"
  }
}
```

Last updated January 22, 2020.