The first part is the image path or URI, while the other is the image label.

6. When preparing the image dataset, it is useful to have a '**None_of_the_above'** image class. This class will contain random images that do not belong to any of the predicted classes. Adding this class can have an overall effect on the model accuracy.

7. Clone the GitHub book repository to the Notebook instance.

8. Navigate to the folder chapter and copy the image files to the GCS bucket.

```
gsutil cp -r cereal_photos gs://quantum-ally-219323-vcm
```

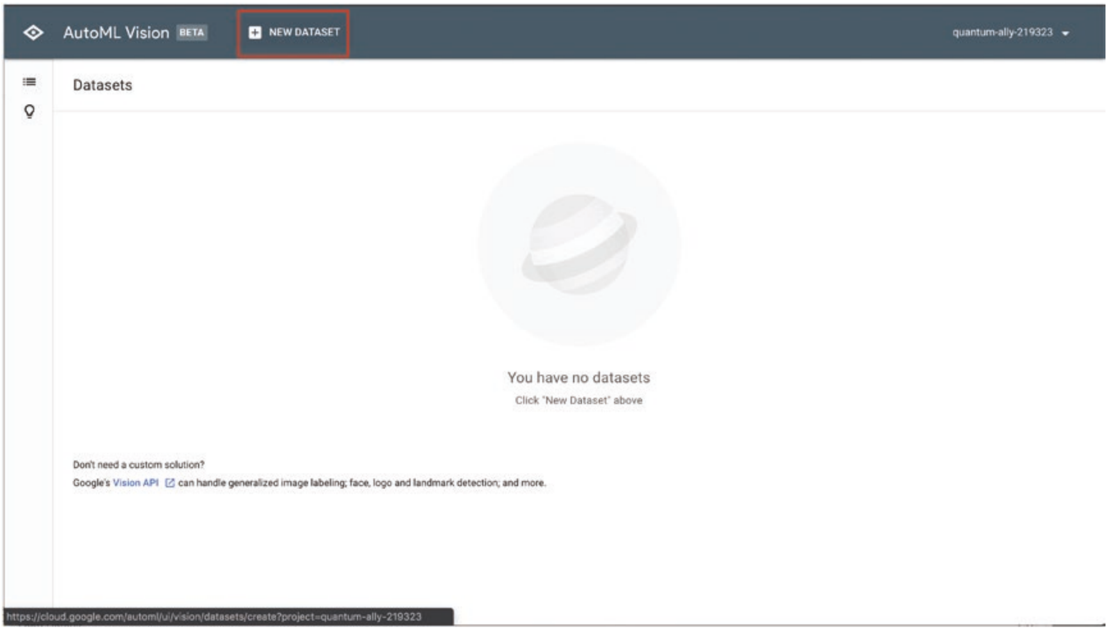
9. Copy the CSV data file containing the image paths and their labels to the GCS bucket.

```
gsutil cp data.csv gs://quantum-ally-219323-vcm/cereal_photos/
```

Building Custom Image Models on Cloud AutoML Vision

In AutoML for Cloud Vision, a dataset contains the images that will be used in building the classifier and their corresponding labels. This section will walk through creating a dataset and building a custom image model on AutoML Vision.

1. From the Cloud AutoML Vision Dashboard, click **NEW DATASET** as shown in Figure 42-8.



You have no datasets

Click "New Dataset" above

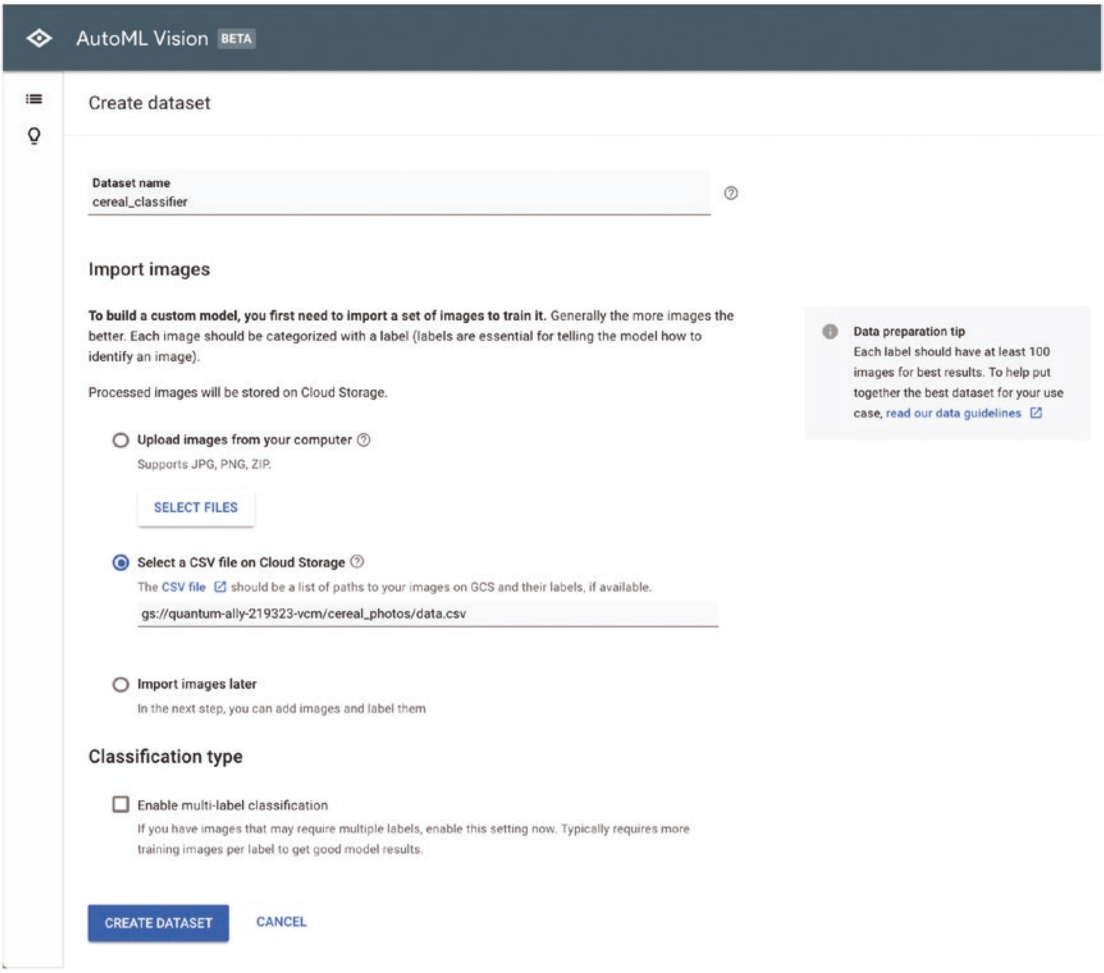
Don't need a custom solution?

Google's Vision API can handle generalized image labeling; face, logo and landmark detection; and more.

https://cloud.google.com/automl/ui/vision/datasets/create?project=quantum-ally-219323

Figure 42-8. *New Dataset on AutoML Vision*

2. To create a Dataset on Cloud AutoML Vision, set the following parameters as shown in Figure 42-9:



Create dataset

Dataset name
cereal_classifier

Import images

To build a custom model, you first need to import a set of images to train it. Generally the more images the better. Each image should be categorized with a label (labels are essential for telling the model how to identify an image).

Processed images will be stored on Cloud Storage.

○ Upload images from your computer
Supports JPG, PNG, ZIP.

● Select a CSV file on Cloud Storage
The CSV file should be a list of paths to your images on GCS and their labels, if available.
gs://quantum-ally-219323-vcm/cereal_photos/data.csv

○ Import images later
In the next step, you can add images and label them

Data preparation tip
Each label should have at least 100 images for best results. To help put together the best dataset for your use case, read our data guidelines

Classification type

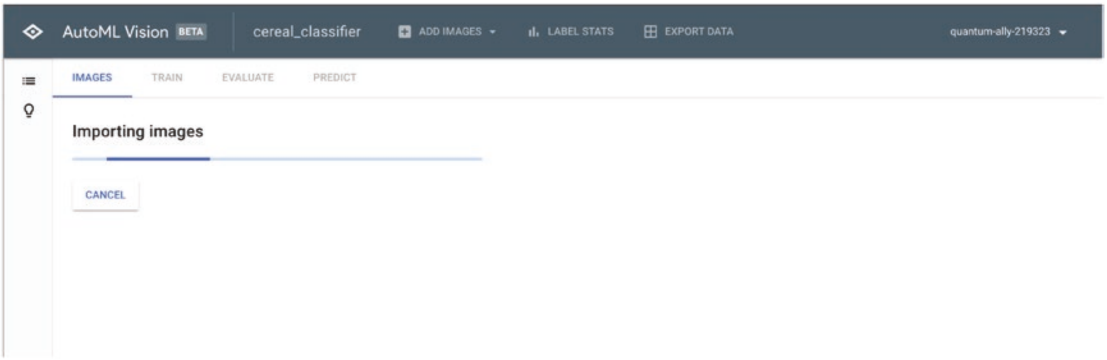
□ Enable multi-label classification
If you have images that may require multiple labels, enable this setting now. Typically requires more training images per label to get good model results.

***Figure 42-9.** Create a Dataset on Cloud AutoML Vision*

a. Dataset name: cereal_classifier.

b. Select a CSV file on Cloud Storage (this is the CSV file placed on the bucket created when Cloud AutoML was configured that contains the path to the images): gs://quantum-ally-219323-vcm/cereal_photos/data.csv.

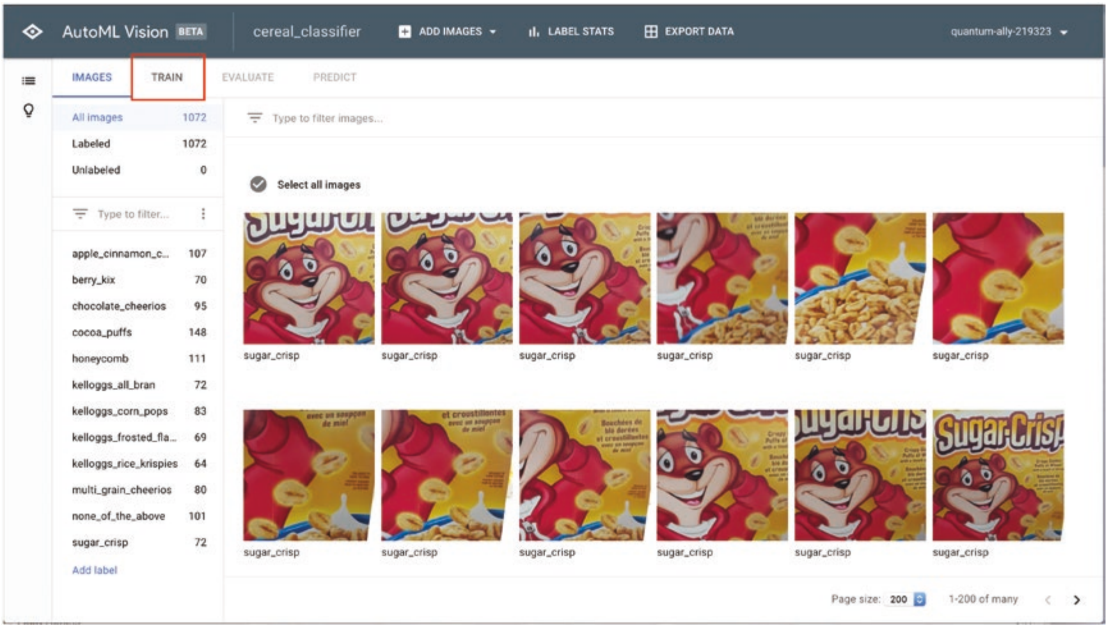
c. Click **CREATE DATASET** to begin importing images (see Figure 42-10).



Importing images

Figure 42-10. *Cloud AutoML Vision: Importing images*

3. After importing the Dataset, click **TRAIN** (see Figure 42-11) to initiate the process of building a custom image recognition model.



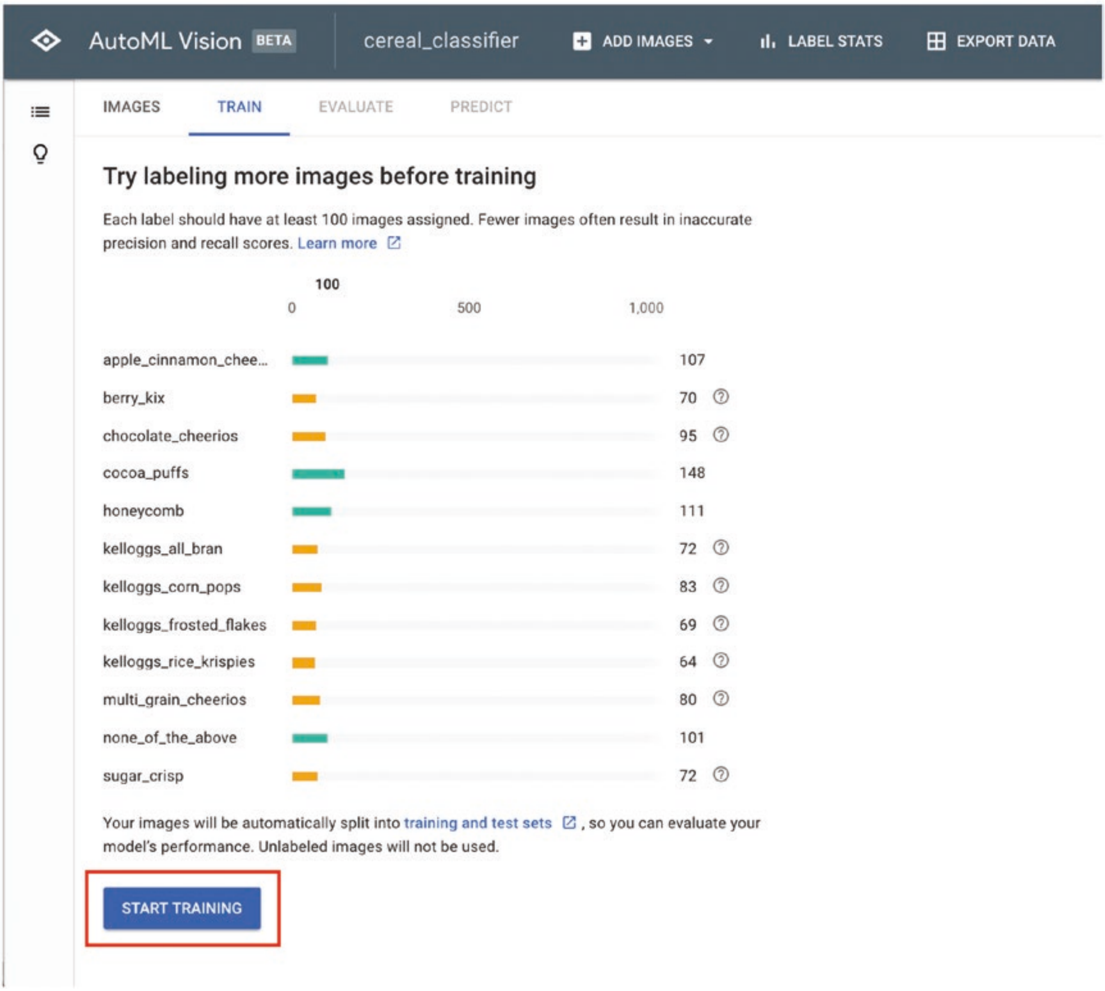
All images	1072
Labeled	1072
Unlabeled	0

apple_cinnamon_c...	107
berry_kix	70
chocolate_cheerios	95
cocoa_puffs	148
honeycomb	111
kelloggs_all_bran	72
kelloggs_corn_pops	83
kelloggs_frosted_fla...	69
kelloggs_rice_krispies	64
multi_grain_cheerios	80
none_of_the_above	101
sugar_crisp	72

Figure 42-11. *Cloud AutoML Vision: Imported images and their labels*

4. In machine learning, more labeled training examples boost the performance of the model. Likewise, when using AutoML, there should be at least 100 training examples for each image class. In the example used in this section, some classes do not have up to

100 examples, so AutoML gives a warning as seen in Figure 42-12. However, for the purposes of this exercise, we will continue with training. Click **START TRAINING**.



Try labeling more images before training

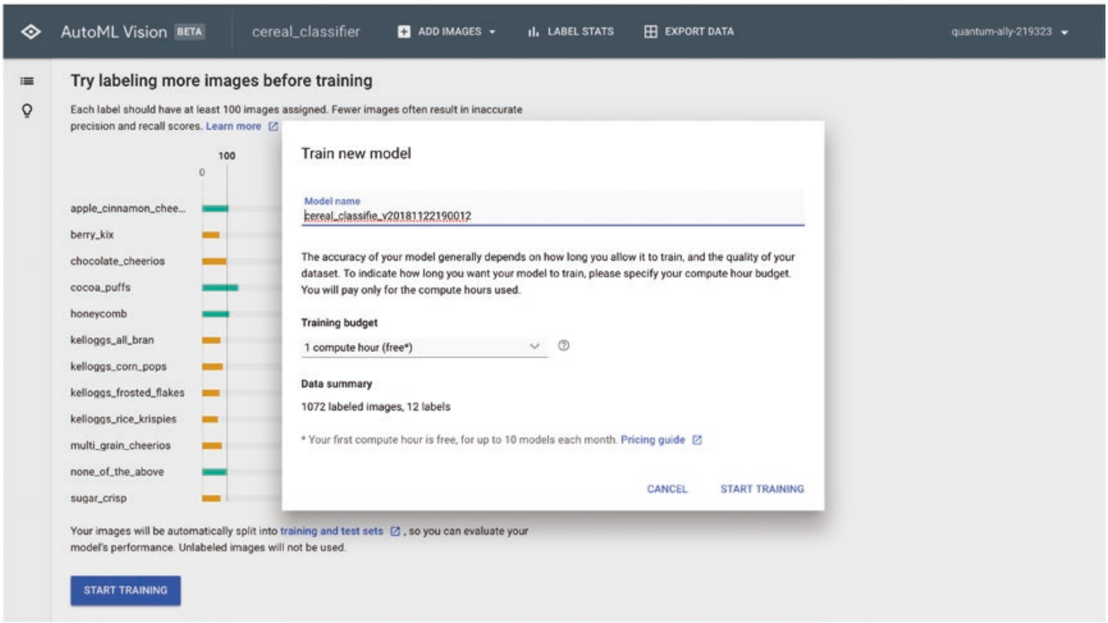
Each label should have at least 100 images assigned. Fewer images often result in inaccurate precision and recall scores. Learn more

apple_cinnamon_chee...	107
berry_kix	70
chocolate_cheerios	95
cocoa_puffs	148
honeycomb	111
kelloggs_all_bran	72
kelloggs_corn_pops	83
kelloggs_frosted_flakes	69
kelloggs_rice_krispies	64
multi_grain_cheerios	80
none_of_the_above	101
sugar_crisp	72

Your images will be automatically split into training and test sets, so you can evaluate your model's performance. Unlabeled images will not be used.

Figure 42-12. *Cloud AutoML Vision requesting for more training examples per image class*

5. Choose how long the model will be trained. More training time might have an effect on the model accuracy, but this may cost more for running on Cloud AutoML's machines (see Figure 42-13). Again, click **START TRAINING** to begin building the model (see Figure 42-14).



Try labeling more images before training

Each label should have at least 100 images assigned. Fewer images often result in inaccurate precision and recall scores. Learn more

Train new model

Model name: cereal_classifie_v20181122190012

The accuracy of your model generally depends on how long you allow it to train, and the quality of your dataset. To indicate how long you want your model to train, please specify your compute hour budget. You will pay only for the compute hours used.

Training budget: 1 compute hour (free*)

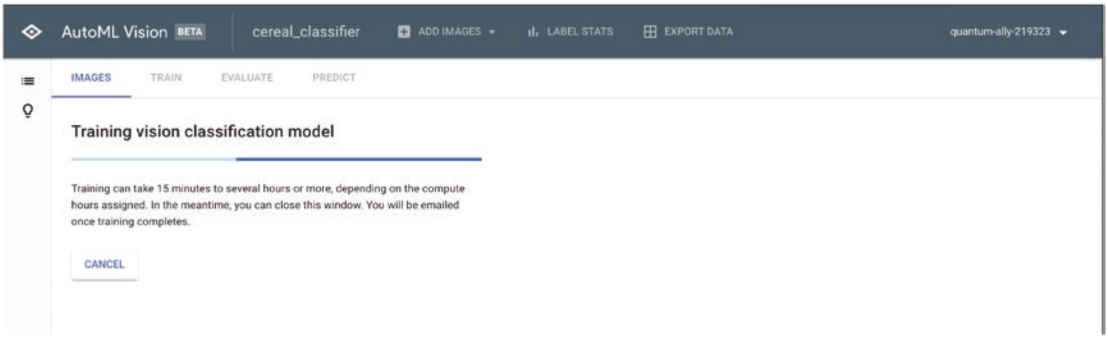
Data summary

1072 labeled images, 12 labels

* Your first compute hour is free, for up to 10 models each month. Pricing guide

Your images will be automatically split into training and test sets, so you can evaluate your model's performance. Unlabeled images will not be used.

Figure 42-13. *Select training budget*

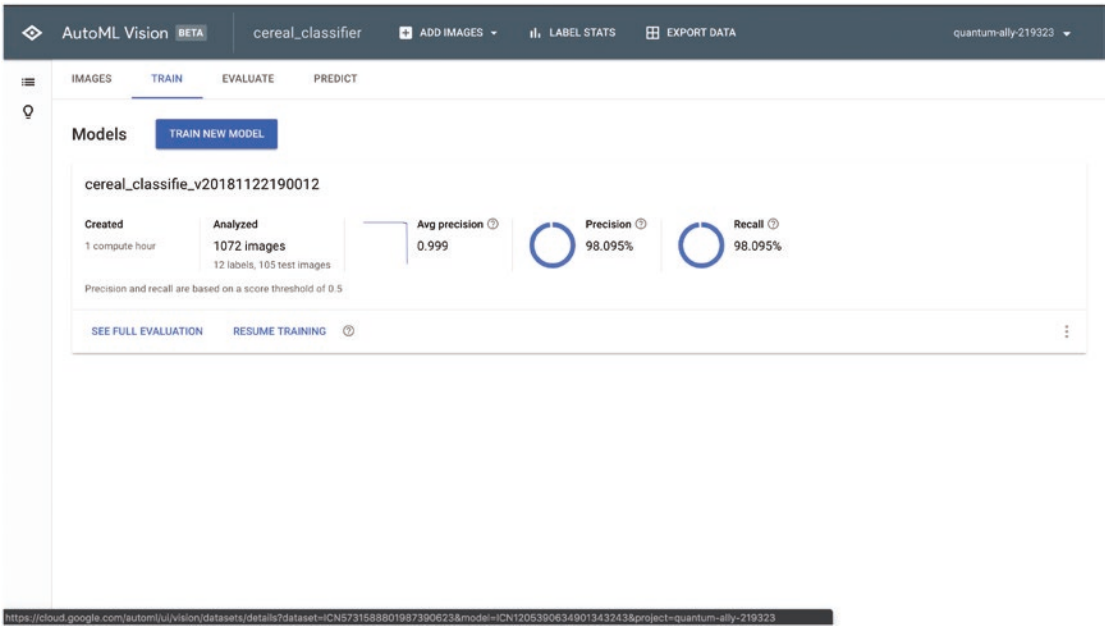


Training vision classification model

Training can take 15 minutes to several hours or more, depending on the compute hours assigned. In the meantime, you can close this window. You will be emailed once training completes.

Figure 42-14. *Training vision model on Cloud AutoML Vision*

6. The training summary is shown in Figure 42-15.



cereal_classifie_v20181122190012

Created: 1 compute hour

Analyzed: 1072 images — 12 labels, 105 test images

Avg precision: 0.999

Precision: 98.095%

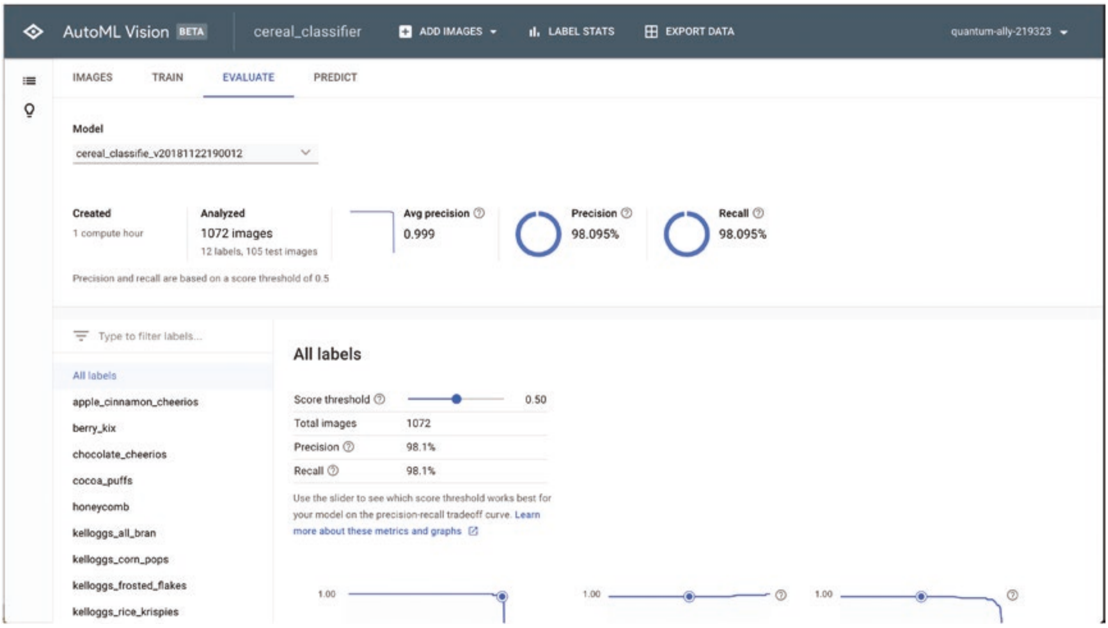
Recall: 98.095%

Precision and recall are based on a score threshold of 0.5

https://cloud.google.com/automl/ui/vision/datasets/details?dataset=ICN5731588801987390623&model=ICN1205390634901343243&project=quantum-ally-219323

Figure 42-15. *Cloud AutoML Vision: Training summary*

7. AutoML Vision uses the set-aside test images to evaluate the quality of the model after training as seen in Figure 42-16. The F1 plot showing the trade-off between precision and recall is shown in Figure 42-17. Also, a visual confusion matrix is provided to further evaluate the model quality (see Figure 42-18).



Model
cereal_classifie_v20181122190012

Created
1 compute hour

Analyzed
1072 images
12 labels, 105 test images

Avg precision
0.999

Precision
98.095%

Recall
98.095%

Precision and recall are based on a score threshold of 0.5

Type to filter labels...

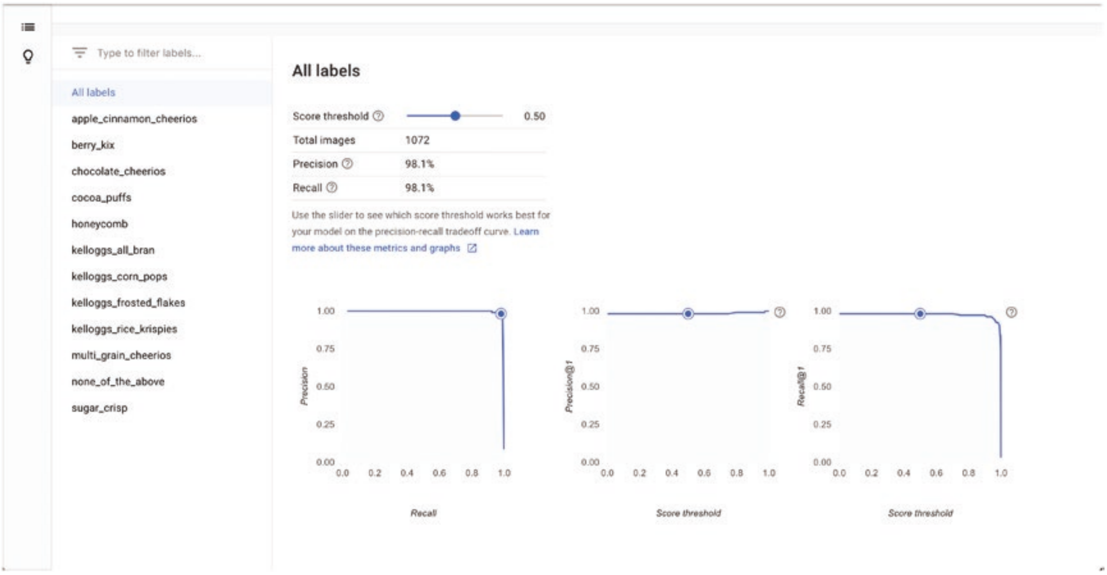
All labels
apple_cinnamon_cheerios
berry_kix
chocolate_cheerios
cocoa_puffs
honeycomb
kelloggs_all_bran
kelloggs_corn_pops
kelloggs_frosted_flakes
kelloggs_rice_krispies

All labels

Score threshold	0.50
Total images	1072
Precision	98.1%
Recall	98.1%

Use the slider to see which score threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs

Figure 42-16. *Cloud AutoML Vision: Model evaluation*



Type to filter labels...

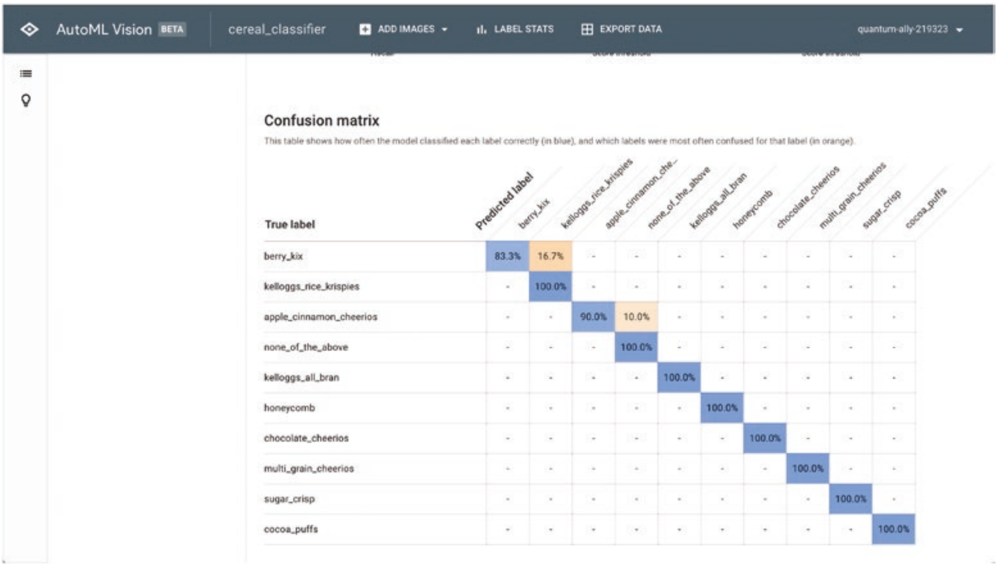
All labels
apple_cinnamon_cheerios
berry_kix
chocolate_cheerios
cocoa_puffs
honeycomb
kelloggs_all_bran
kelloggs_corn_pops
kelloggs_frosted_flakes
kelloggs_rice_krispies
multi_grain_cheerios
none_of_the_above
sugar_crisp

All labels

Score threshold	0.50
Total images	1072
Precision	98.1%
Recall	98.1%

Use the slider to see which score threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs

Figure 42-17. *F1 evaluation matrix on Cloud AutoML Vision*



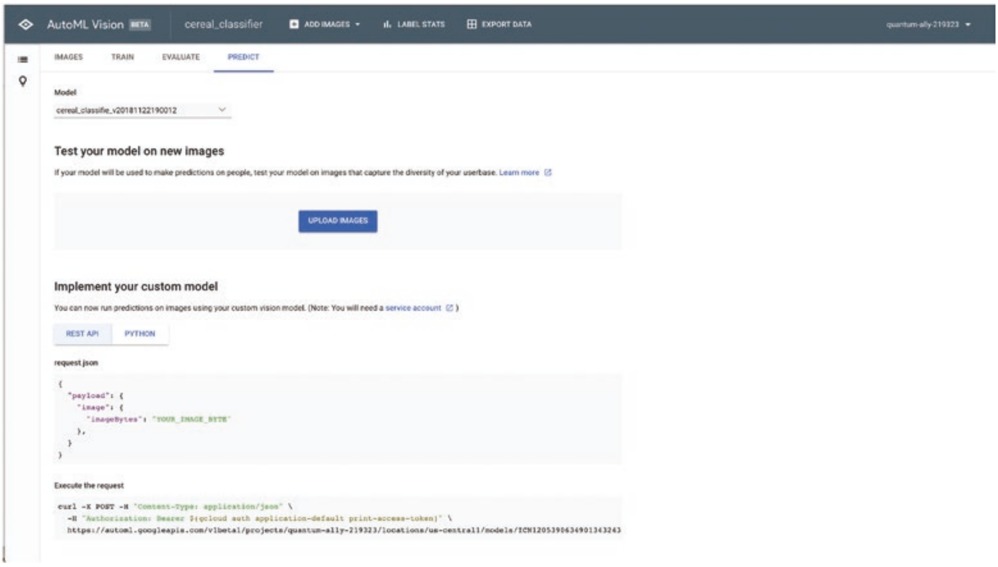
Confusion matrix

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in orange).

True label \ Predicted label	berry_kix	kelloggs_rice_krispies	apple_cinnamon_che...	none_of_the_above	kelloggs_all_bran	honeycomb	chocolate_cheerios	multi_grain_cheerios	sugar_crisp	cocoa_puffs
berry_kix	83.3%	16.7%	-	-	-	-	-	-	-	-
kelloggs_rice_krispies	-	100.0%	-	-	-	-	-	-	-	-
apple_cinnamon_cheerios	-	-	90.0%	10.0%	-	-	-	-	-	-
none_of_the_above	-	-	-	100.0%	-	-	-	-	-	-
kelloggs_all_bran	-	-	-	-	100.0%	-	-	-	-	-
honeycomb	-	-	-	-	-	100.0%	-	-	-	-
chocolate_cheerios	-	-	-	-	-	-	100.0%	-	-	-
multi_grain_cheerios	-	-	-	-	-	-	-	100.0%	-	-
sugar_crisp	-	-	-	-	-	-	-	-	100.0%	-
cocoa_puffs	-	-	-	-	-	-	-	-	-	100.0%

__Figure 42-18.__ Confusion matrix for model evaluation on Cloud AutoML Vision

8. The custom image recognition model is exposed as a REST or Python API for integration into software applications as a prediction service (see Figure 42-19). We can test our model by uploading a sample image for classification as shown in Figure 42-20.



Model: cereal_classifie_v20181122190012

Test your model on new images

If your model will be used to make predictions on people, test your model on images that capture the diversity of your userbase. Learn more

Implement your custom model

You can now run predictions on images using your custom vision model. (Note: You will need a service account)

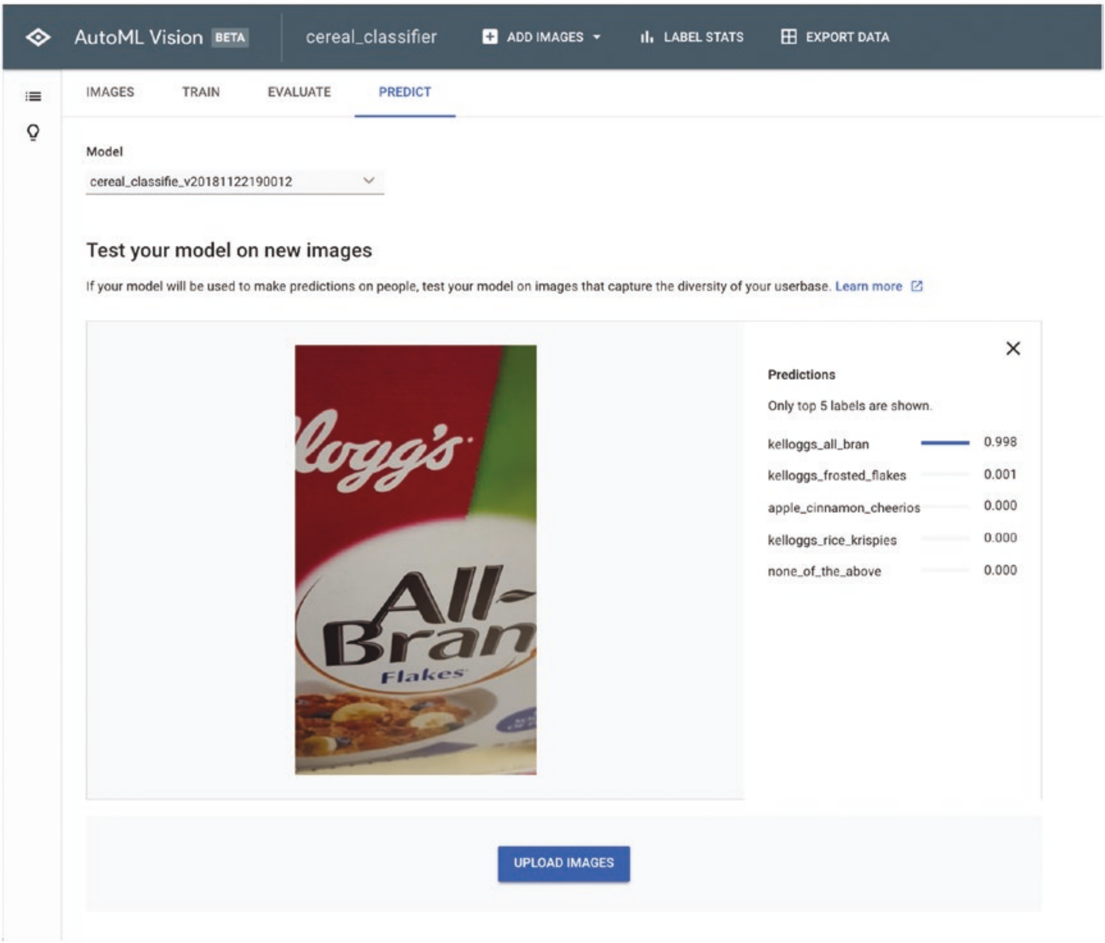
request json

```
{
  "payload": {
    "image": {
      "imageBytes": "YOUR_IMAGE_BYTE"
    },
  }
}
```

Execute the request

```
curl -X POST -H "Content-Type: application/json" \
  -H "Authorization: Bearer $(gcloud auth application-default print-access-token)" \
  https://automl.googleapis.com/v1beta1/projects/quantum-ally-219323/locations/us-central1/models/ICN1205390634901343243
```

__Figure 42-19.__ Cloud AutoML Vision: Model as a prediction service



Model

cereal_classifie_v20181122190012

Test your model on new images

If your model will be used to make predictions on people, test your model on images that capture the diversity of your userbase. Learn more

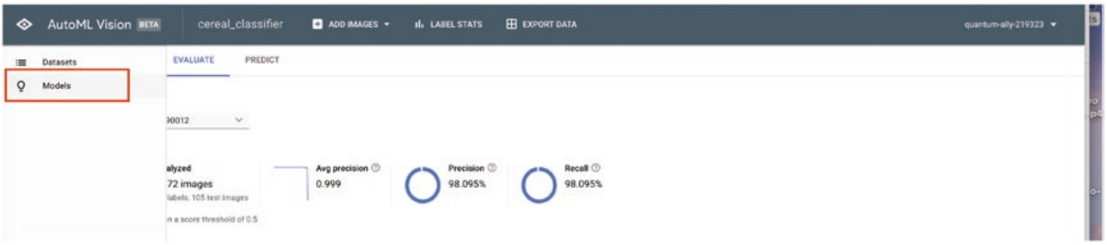
Predictions

Only top 5 labels are shown.

kelloggs_all_bran	0.998
kelloggs_frosted_flakes	0.001
apple_cinnamon_cheerios	0.000
kelloggs_rice_krispies	0.000
none_of_the_above	0.000

Figure 42-20. *Test prediction service on Cloud AutoML Vision*

9. To delete a model, click the triple dash and select Models to navigate to the Models Dashboard (see Figure 42-21). At the side of the model, click the triple dot and select Delete model (see Figure 42-22). Confirm deletion as shown in Figure 42-23. Note, however, that API calls affiliated with a deleted model will cease to be operational.



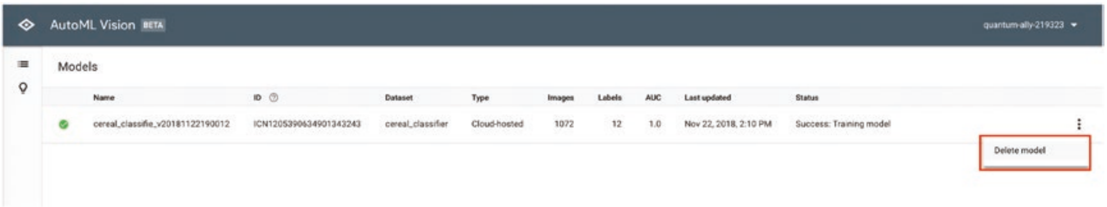
alyzed
72 images
labels, 105 test images
n a score threshold of 0.5

Avg precision 0.999

Precision 98.095%

Recall 98.095%

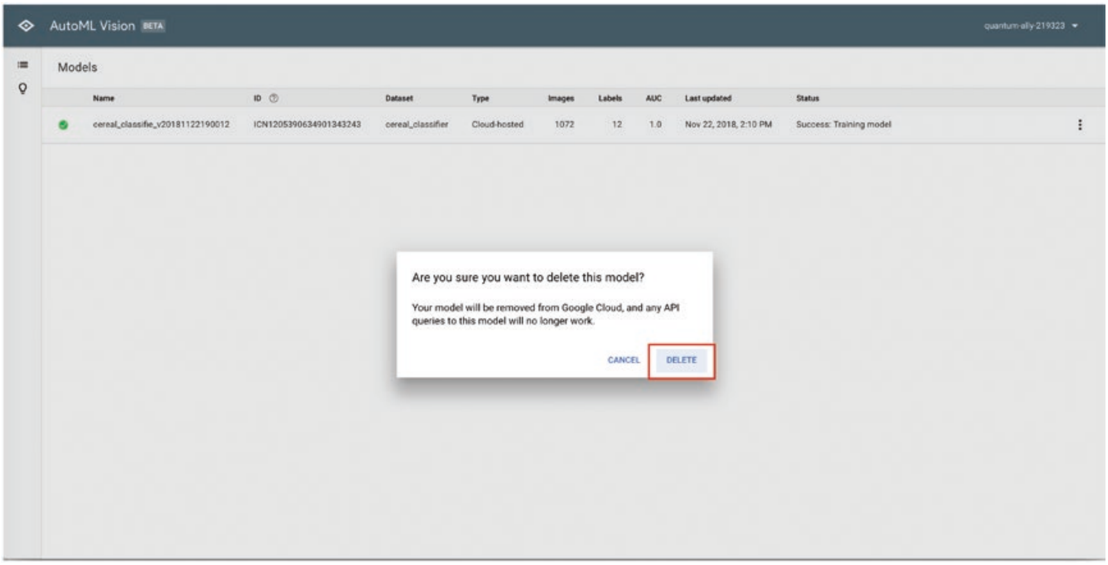
Figure 42-21. Return to Models dashboard



Models

Name	ID	Dataset	Type	Images	Labels	AUC	Last updated	Status
cereal_classifie_v20181122190012	ICN1205390634901343243	cereal_classifier	Cloud-hosted	1072	12	1.0	Nov 22, 2018, 2:10 PM	Success: Training model

Figure 42-22. Select model to delete



Models

Name	ID	Dataset	Type	Images	Labels	AUC	Last updated	Status
cereal_classifie_v20181122190012	ICN1205390634901343243	cereal_classifier	Cloud-hosted	1072	12	1.0	Nov 22, 2018, 2:10 PM	Success: Training model

Are you sure you want to delete this model?

Your model will be removed from Google Cloud, and any API queries to this model will no longer work.

Figure 42-23. Delete a model on Cloud AutoML Vision

This chapter covered building and deploying custom image classification models using Google AutoML Cloud Vision. In the next chapter, we will discover how to build and deploy custom text classification models with Google Cloud AutoML for natural language processing.

CHAPTER 43

Google AutoML: Cloud Natural Language Processing

This chapter will build a language toxicity classification model to classify and recognize toxic and non-toxic or clean phrases using Google Cloud AutoML for natural language processing (NLP). The data used in this project is from the Toxic Comment Classification Challenge on Kaggle by Jigsaw and Google. The data is modified to have a sample of 16,000 toxic and 16,000 non-toxic words as inputs to build the model on AutoML NLP.

Enable AutoML NLP on GCP

The following steps will enable AutoML NLP on GCP:

1. Click the triple dash in the top-left corner of the interface and select **Natural Language** under the category ARTIFICIAL INTELLIGENCE as shown in Figure 43-1.

E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*,
https://doi.org/10.1007/978-1-4842-4470-8_43