Quotas & Limits

This document contains current restrictions and usage quotas on use of Cloud AutoML Vision Object Detection. This page will be updated to reflect any changes to these restrictions and usage quotas.

These quotas apply to each Cloud AutoML Vision Object Detection project and are shared across all applications and IP addresses using a given project.

You can edit your quotas up to their maximum values by selecting **Edit Quotas** from the **Quotas** page of the Google Cloud Platform Dashboard
  (https://console.cloud.google.com/project/_/quotas?service=automl.googleapis.com). Edit the quotas for the **Cloud AutoML API**. To request an increase above the maximum quota, edit your quota with your requested increase and justification and submit your update. You will be notified when your request has been received. You might be contacted for more information regarding your request. After your request has been reviewed, you will be notified whether it has been approved or denied.

Cloud AutoML Vision Object Detection has quotas associated with creating models, human labeling, and making requests to the models.

Modeling quotas

Activity	Default quota
Concurrent model training (creation) requests	5

Request quotas

The current usage quotas for AutoML Vision Object Detection for non-prediction requests is 600 per minute per project. It applies collectively to all deployed models.

Online prediction request quotas

The number of online prediction per minute for Cloud AutoML Vision Object Detection are based on how many nodes are deployed. As is explained on the pricing (https://cloud.google.com/vision/automl/pricing) page, many customers find that with one node hour they can serve maximum 1.5 QPS. You can adjust the number of nodes when you deploy your model. However, the total number of nodes across all the deployed models in the project is subject to the quota below.

Activity	Default quota
The number of deployed model nodes	10

Offline batch prediction requests

The current usage quotas for Cloud AutoML Vision Object Detection are 5 *concurrent* batch prediction requests per project. These quotas apply collectively to all deployed models in a project.

Limits

Type of limit	Value
Image file size	Maximum: 30MB
Images per dataset	Maximum: 1,000,000
Labels per dataset	Minimum: 1 Maximum: 100
Images per label	Minimum: 10 Recommended: 1000
Batch input CSV file size	Maximum: 100MB

Type of limit	Value
Number of images in batch input	Maximum: 300,000

Review quotas on Google Cloud Console

There are two main ways to view your current quota limits in the Google Cloud Console
 (https://console.cloud.google.com/):

- Using the Quotas (https://console.cloud.google.com/quotas?project=_) page, which gives you a
  list of all your project's quota usage and limits.

- Using the console (https://console.cloud.google.com/apis/dashboard), which gives you quota
  information for a particular API, including resource usage over time.

Locate specific operation quotas in the Quotas (https://console.cloud.google.com/quotas?project=_)
page by first selecting `Cloud AutoML API` from the **Service** menu. With **Service:** `Cloud AutoML API` selected you can then select the appropriate **Metric**.
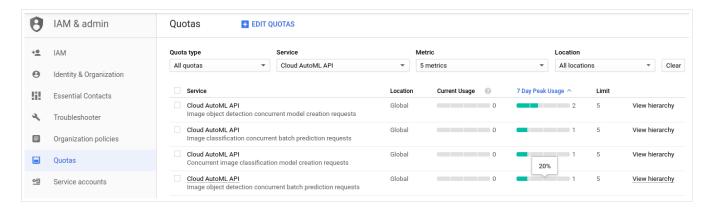
Examples:

Operation description	Method name
Image classification: Simultaneous model training	projects.locations.models.create (https://cloud.google.com/automl/docs/reference/r...
Image classification: Online prediction	projects.locations.models.predict (https://cloud.google.com/automl/docs/reference/...
Object detection:	projects.locations.models.create (https://cloud.google.com/automl/docs/reference/r...

Simultaneous model training	
Object detection: Simultaneous offline batch prediction	`projects.locations.models.batchPredict` (https://cloud.google.com/automl/docs/refe

Quotas page:

Last updated December 4, 2019.