Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 126 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 126
Topic #: 1
[All Professional Data Engineer Questions]

You work for a manufacturing company that sources up to 750 different components, each from a different supplier. You've collected a labeled dataset that has on average 1000 examples for each unique component. Your team wants to implement an app to help warehouse workers recognize incoming components based on a photo of the component. You want to implement the first working version of this app (as Proof-Of-Concept) within a few working days. What should you do?

A. Use Cloud Vision AutoML with the existing dataset.

B. Use Cloud Vision AutoML, but reduce your dataset twice.

C. Use Cloud Vision API by providing custom labels as recognition hints.

D. Train your own image recognition model leveraging transfer learning techniques.

Show Suggested Answer

by [deleted] at *March 22, 2020, 10:52 a.m.*

Comments

Type your comment...

Submit

Callumr Highly Voted 4 years, 4 months ago

B - You only need a PoC and it has be done quickly

upvoted 54 times

[Removed] Highly Voted 4 years, 7 months ago

Correct - A

upvoted 20 times

grshankar9 Most Recent 3 months, 2 weeks ago

Selected Answer: A

The key difference between Google Cloud Vision AutoML and Cloud Vision API is that Cloud Vision API provides pre-trained models for basic image analysis tasks like object detection and labeling, while Cloud Vision AutoML allows you to train custom machine learning models to identify specific objects or concepts within images that are unique to your dataset, requiring you to provide labeled training data.

upvoted 1 times

josech 5 months, 1 week ago

Selected Answer: A

AutoML Vision is deprecated since march 31, 2024. The question will refer to Vertex AI AutoML. And as bet practice, the minimum dataset size for each label is 1000. So, with an updated question, the answer would be A.

upvoted 3 times

CGS22 7 months ago

Selected Answer: A

A. Use Cloud Vision AutoML with the existing dataset.

Here's why this is the most suitable option:

Speed and Ease: AutoML simplifies model building. You simply upload your labeled images, and AutoML takes care of model selection, training, and evaluation.
Existing Dataset Sufficiency: Your dataset (750 components x 1000 images each) is a decent starting point for AutoML, allowing you to quickly test its effectiveness.
Minimal Custom Development: AutoML's out-of-the-box deployment options let you integrate the model into your app without extensive coding.

upvoted 1 times

saado9 1 year, 1 month ago

Selected Answer: B

Option B is the fastest way to train a model that can be used to recognize the 750 different components.

upvoted 1 times

musumusu 1 year, 8 months ago

Whats wrong with C, its fast, cheap and add your 750 labels which is not big work.
AutoML is good to train on big dataset and costly as compared to APIs

upvoted 2 times

forepick 1 year, 5 months ago

Adding custom labels to Vision API is done by training an AutoML model! That's the formal recommendation. And you don't need a big dataset for AutoML as it uses transfer learning.

upvoted 4 times

knith66 1 year, 3 months ago

it is a labeled dataset and why do you need to label it once again? So no C

upvoted 1 times

techtitan 1 year, 9 months ago

A - https://cloud.google.com/vertex-ai/docs/beginner/beginners-guide Target at least 1000 examples per target

upvoted 8 times

techtitan 1 year, 9 months ago

The quick POC part can be achieved by using Auto ML instead of creating and training your own model

upvoted 1 times

odacir 1 year, 11 months ago

Selected Answer: A

First I think in Vision API, but that is a pre-trained AI, will not recognize my labels, so because you have 1000 samples per item, AUTO ML is perfect. B cannot be because have not sensed to reduce your dataset if you have the recommended number of info.
https://cloud.google.com/vision/automl/docs/beginners-guide#include_enough_labeled_examples_in_each_category
The bare minimum required by AutoML Vision training is 100 image examples per category/label. The likelihood of successfully recognizing a label goes up with the number of high quality examples for each; in general, the more labeled data

you can bring to the training process, the better your model will be. Target at least 1000 examples per label.

upvoted 8 times

AzureDP900 1 year, 10 months ago

A is correct

upvoted 2 times

zellck 1 year, 11 months ago

Selected Answer: A

A is the answer.

https://cloud.google.com/vision/automl/docs/beginners-guide#include_enough_labeled_examples_in_each_category
The bare minimum required by AutoML Vision training is 100 image examples per category/label. The likelihood of successfully recognizing a label goes up with the number of high quality examples for each; in general, the more labeled data you can bring to the training process, the better your model will be. Target at least 1000 examples per label.

upvoted 4 times

ga8our 1 year, 5 months ago

So how are you going to test that the model was able to adequately learn from the sample? The point of splitting a dataset is to train the model on one part of the data (say 80%), and then test it on the other part (20%). If your model is able to predict the outcome of (most of) the sample points in your test dataset, you can be confident that it will work well on future data. Without a test data set, however, you have no such feedback. Therefore, the answer is B.

upvoted 2 times

NewDE2023 1 year, 3 months ago

I believe that the ideal would be to reduce the number of components for the POC and preserve the number of examples, so my answer is A.

upvoted 1 times

odacir 1 year, 11 months ago

Agreed !

upvoted 1 times

gudiking 1 year, 11 months ago

A - https://cloud.google.com/vision/automl/docs/beginners-guide#include_enough_labeled_examples_in_each_category

upvoted 1 times

MarielaYBird 1 year, 11 months ago

Selected Answer: B

Based on this:
"As a rule of thumb, we recommend to have at least 100 training samples per class if you have distinctive and few classes, and more than 200 training samples if the classes are more nuanced and you have more than 50 different classes"

750 different components = more than 50 different classes. That means we need more than 200 training samples. If we used 250 training samples out of the 1000 samples and multiply it to 750 different classes we get a total of 187,500 which is the equivalent of reducing the dataset twice.

https://cloud.google.com/vision/automl/object-detection/docs/prepare#how_big_does_the_dataset_need_to_be

upvoted 5 times

josrojgra 2 years ago

Selected Answer: A

I choose A because on the vertex AI documentation (https://cloud.google.com/vertex-ai/docs/image-data/classification/prepare-data), on the best practices of preparing data for image recognition recommend this: We recommend about 1000 training images per label. The minimum per label is 10. In general, it takes more examples per label to train models with multiple labels per image, and resulting scores are harder to interpret.

I know that is PoC, but if you do it without enough accuracy, you maybe discard the solution because it isn't fit for your requirements. So is better to do it with enough data to be sure that the model is or not accuracy enough with this data, because you maybe haven't enough accuracy and the problem is the quality of the data and not the amount of it.

upvoted 3 times

John_Pongthorn 2 years, 1 month ago

Selected Answer: A

https://cloud.google.com/vision/automl/docs/beginners-guide#include_enough_labeled_examples_in_each_category

The bare minimum required by AutoML Vision training is 100 image examples per category/label. The likelihood of successfully recognizing a label goes up with the number of high quality examples for each; in general, the more labeled data you can bring to the training process, the better your model will be. Target at least 1000 examples per label.

upvoted 3 times

John_Pongthorn 2 years, 1 month ago

The more labels, the more accurate the result.

upvoted 1 times

changsu 2 years, 1 month ago

Selected Answer: B

750*1000 are a lot.

upvoted 1 times

ducc 2 years, 2 months ago

Selected Answer: A

It is labeled, so A is correct

upvoted 1 times

civilizador 2 years, 2 months ago

It's A.
https://cloud.google.com/vision/automl/docs/beginners-guide#data_preparation

The bare minimum required by AutoML Vision training is 100 image examples per category/label. The likelihood of successfully recognizing a label goes up with the number of high quality examples for each; in general, the more labeled data you can bring to the training process, the better your model will be. Target at least 1000 examples per label.

upvoted 5 times

civilizador 2 years, 2 months ago

So even for POC better to use 1000 . There would be no significant time differences anyway between 500 and 1000

upvoted 1 times

Load full discussion...