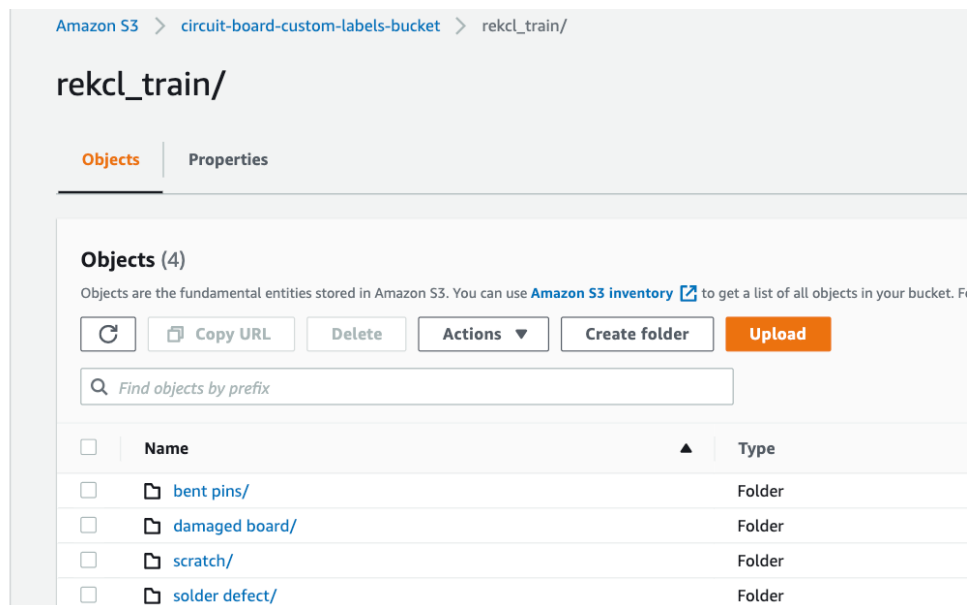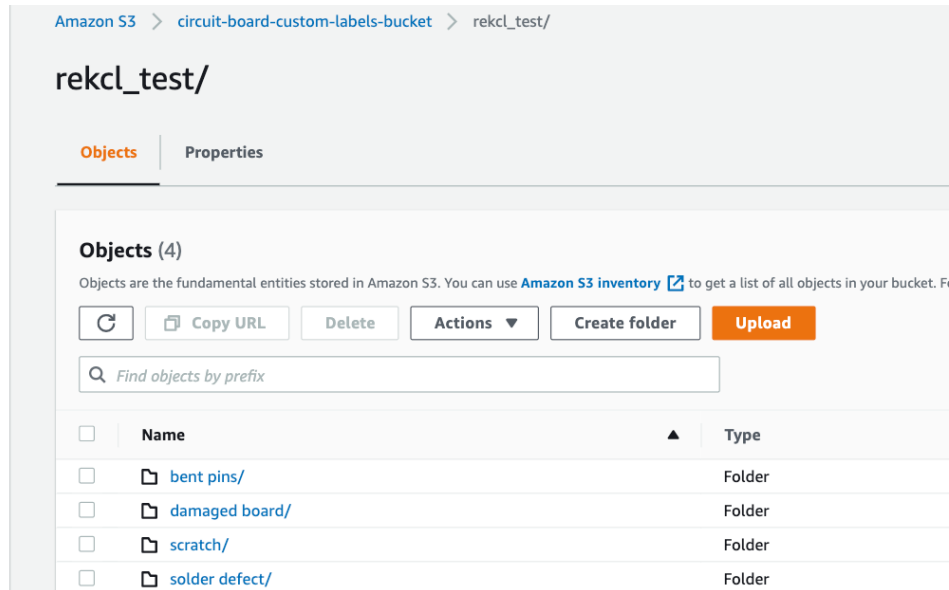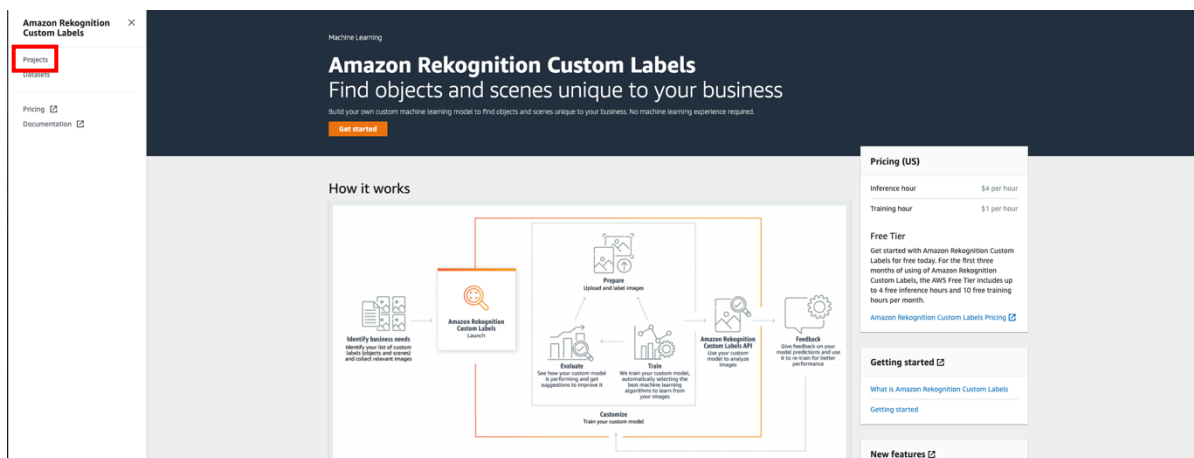# Amazon Rekognition Custom Labels Model Training

Moving on to the next step, in this section we will be using Amazon Rekognition Custom Labels multi-label classification, to train the model on different types of anomalies for circuit boards.

1. **Preparing the dataset:** In Amazon Rekognition Custom Labels we will train the model on the various types of defects such as solder defect, bent pins, damaged board, and scratches.

    a. Open the S3 service and create a new bucket in the same region as Amazon Rekognition and Amazon Lookout for Vision services to load the training images. The following figure shows the Amazon S3 folder structure to train and test multi-class labels.

b. On the Amazon Rekognition page, select **Use Custom Labels** under
    Custom Labels. As seen in the figure below, on the **Custom Labels** page,
    select **Projects**, choose **New Project**, give the project a unique name, and
    select **Create**. If you are using Custom Labels for the first time, you will be
    asked to to create a default S3 bucket to host your projects. **Select OK**.

2. **Create the dataset:**

    a. We will be creating 2 datasets: a **training dataset** and a **testing dataset**.
        On the Amazon Rekognition Custom Labels page, select **Create dataset**
        and give it a unique name for training. Fill out the options as highlighted
        in the figure below. Amazon Rekognition will provide an S3 bucket policy
        to use for the S3 bucket with the dataset. Copy the bucket policy, edit the
        S3 bucket policy for the dataset bucket and paste and save the policy.
        Choose **Submit** for the Amazon Rekognition dataset creation form.

Dataset details

Dataset name

circuit-board-custom-labels-train

The dataset name can't be more than 50 characters. It needs to be a valid S3 path with no spaces.

Image location

Import images from one of the options below.

○ Import images labeled by Amazon SageMaker Ground Truth
Provide the location of your .manifest file. If you've labeled datasets in a different format, convert them to a .manifest format.

● Import images from Amazon S3 bucket
Use images from an existing S3 bucket by entering the S3 folder location below. You have the option to automatically add labels based on your folder names.

○ Copy an existing Amazon Rekognition Custom Labels dataset
Use an existing dataset as a starting point for your new dataset. Your original dataset will remain unchanged.

○ Upload images from your computer
Add images by dragging and dropping them into your dataset gallery view. You're limited to uploading 30 images at one time.

S3 folder location

S3 bucket location of your images.

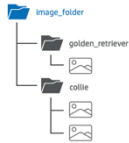s3://circuit-board-custom-labels-bucket/rekcl_train/

Supported image formats: JPG, PNG. Maximum images per dataset: 250,000. Maximum image size: 15 MB, Minimum size (px): 64 x 64. Maximum size (px): 4096 x 4096.

For best results, we recommend uploading images from folders within the S3 bucket created for you during first-time setup.

Automatic labeling

If you've organized the images in your S3 bucket by folder name (/Golden-Retriever/01.jpeg), Amazon Rekognition Custom Labels can automatically label these images.

☑ Automatically attach a label to my images based on the folder they're stored in.

📁 image_folder

    📁 golden_retriever

        🖼

    📁 collie

        🖼

        🖼

Make sure that your S3 bucket is correctly configured

You've specified an external S3 bucket: **circuit-board-custom-labels-bucket**.
To use the images in this bucket, **copy the policy below** (to copy, choose the preceding link text). **Paste the policy into the "Bucket Policy" section of circuit-board-custom-labels-bucket.**
If **circuit-board-custom-labels-bucket** already has these permissions, choose **Submit**.
⚠ If you don't apply this policy, you won't be able to train a model from this dataset.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AWSRekognitionS3AclBucketRead20191011",
            "Effect": "Allow",
            "Principal": {
                "Service": "rekognition.amazonaws.com"
            },
            "Action": [
                "s3:GetBucketAcl",
                "s3:GetBucketLocation"
            ],
            "Resource": "arn:aws:s3:::circuit-board-custom-labels-bucket"
        },
        {
            "Sid": "AWSRekognitionS3GetBucket20191011",
            "Effect": "Allow",
            "Principal": {
                "Service": "rekognition.amazonaws.com"
            },
            "Action": [
                "s3:GetObject",
                "s3:GetObjectAcl",
                "s3:GetObjectVersion",
                "s3:GetObjectTagging"
            ],
            "Resource": "arn:aws:s3:::circuit-board-custom-labels-bucket/*"
        },
        {
            "Sid": "AWSRekognitionS3ACLBucketWrite20191011",
            "Effect": "Allow",
            "Principal": {
                "Service": "rekognition.amazonaws.com"
            },
            "Action": "s3:GetBucketAcl",
            "Resource": "arn:aws:s3:::circuit-board-custom-labels-bucket"
        },
        {
            "Sid": "AWSRekognitionS3PutObject20191011",
            "Effect": "Allow",
            "Principal": {
                "Service": "rekognition.amazonaws.com"
            },
            "Action": "s3:PutObject",
            "Resource": "arn:aws:s3:::circuit-board-custom-labels-bucket/*",
            "Condition": {
                "StringEquals": {
                    "s3:x-amz-acl": "bucket-owner-full-control"
                }
            }
        }
    ]
}
```

    b.   Repeat the same steps for the testing dataset as shown in the figure below.

Create dataset
You can create a dataset by importing images and labeling them, or by importing an Amazon SageMaker Ground Truth manifest file. To train a custom model, you need a labeled dataset that accurately reflects the objects and scenes you want to find.

Dataset details

Dataset name

circuit-board-custom-labels-test

The dataset name can't be more than 50 characters. It needs to be a valid S3 path with no spaces.

Image location
Import images from one of the options below.

○ Import images labeled by Amazon SageMaker Ground Truth
Provide the location of your .manifest file. If you've labeled datasets in a different format, convert them to a .manifest format.

● Import images from Amazon S3 bucket
Use images from an existing S3 bucket by entering the S3 folder location below. You have the option to automatically add labels based on your folder names.

○ Copy an existing Amazon Rekognition Custom Labels dataset
Use an existing dataset as a starting point for your new dataset. Your original dataset will remain unchanged.

○ Upload images from your computer
Add images by dragging and dropping them into your dataset gallery view. You're limited to uploading 30 images at one time.

S3 folder location
S3 bucket location of your images.

s3://circuit-board-custom-labels-bucket/rekcl_test/

Supported image formats: JPG, PNG. Maximum images per dataset: 250,000. Maximum image size: 15 MB. Minimum size (px): 64 x 64. Maximum size (px): 4096 x 4096.
For best results, we recommend uploading images from folders within the S3 bucket created for you during first-time setup.

Automatic labeling
If you've organized the images in your S3 bucket by folder name (/Golden-Retriever/01.jpeg), Amazon Rekognition Custom Labels can automatically label these images.
☑ Automatically attach a label to my images based on the folder they're stored in.

image_folder
└ golden_retriever
  └ 🖼
└ collie
  └ 🖼

c. Verify that there 33 labeled images in the training dataset and 18 labeled images in the test dataset as seen below. You will see that there are more images for the **solder defect** label vs all the other defects for train and test. Since the number of images for **damaged board**, **bent pins** and **scratch** is low, the model may not perform well for those labels.
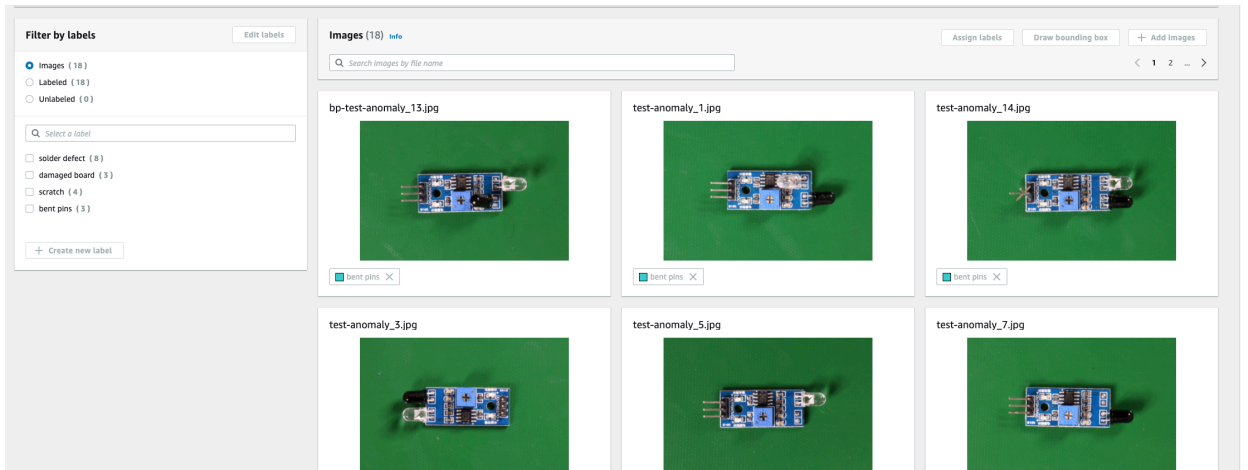
3. Train the Amazon Rekognition Custom Labels model: Once the dataset is created, we will train the model.

    a. Choose **Train Model**. Select the project for anomaly identification that was created under **Choose project**. Select the training and test datasets created in the earlier step. Select **Train** to start training the model
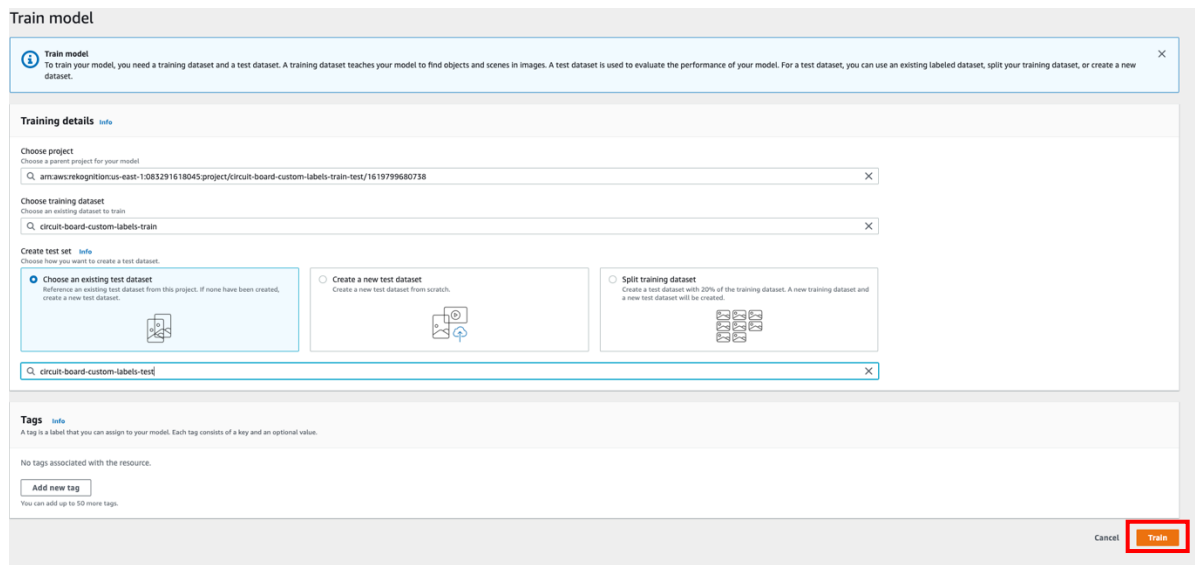
    b. The model training on an average could take a couple of hours to complete. After the model is trained, check the quality of the model by observing its evaluation results and decide whether the model can be used for inference or needs further training.