

Amazon SageMaker Ground Truth

1. Create workforce

Amazon SageMaker

Amazon SageMaker Studio

대시보드

검색

Ground Truth

- 레이블링 작업
- 데이터 세트에 레이블 지정
- 레이블링 인력

노트북

- 노트북 인스턴스
- 수명 주기 구성
- Git 리포지토리

처리 중

- 작업 처리 중

훈련

- 알고리즘
- 훈련 작업
- 하이퍼파라미터 튜닝 작업

추론

- 컴파일 작업
- 모델 패키지
- 모델

의견 한국어

Amazon SageMaker > 레이블 지정 작업 인력 > 프라이빗 팀 만들기

프라이빗 팀 만들기

Private team creation

Create a private team with AWS Cognito
Create a private work team by sending email invitations to new workers or importing workers from existing Amazon Cognito user groups.

Create a private team with OpenID Connect (OIDC)
Create a private work team with your own identity provider (IdP). Your IdP must support OIDC user groups.

팀 세부 정보

팀 이름
작업팀에 알아보기 쉬운 이름을 지정합니다. 이 이름은 나중에 변경할 수 없습니다.
팀 이름 입력

최대 63자의 알파벳 문자 및 숫자, 하이픈을 포함할 수 있지만 공백은 포함할 수 없습니다. AWS 리전의 계정에서 고유해야 합니다.

작업자 추가 정보
작업자 이메일 주소를 추가하거나, 기존 Amazon Cognito 사용자 그룹에서 작업자를 가져와서 프라이빗 작업팀에 작업자를 추가합니다.

이메일로 새 작업자 초대
 기존 Amazon Cognito 사용자 그룹에서 작업자 가져오기

이메일 주소

초대장 미리 보기

© 2008 - 2020, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

- SageMaker의 Labeling workforce에서 Create private team을 선택해 줍니다.
- 각 항목들을 입력하여 private 팀을 생성해 줍니다.

프라이빗 작업 인력 요약 정보
작업 인력을 Amazon Cognito 사용자 풀에서 지원합니다. 사용자 풀을 사용하여 작업자는 Amazon Cognito를 통해 레이블 지정 포털에 로그인할 수 있습니다.

Amazon Cognito 사용자 풀
us-east-1_8OqZg77p8 []

액 클라이언트
1d3272645r7ej94hr18uaedque []

레이블 지정 포털 로그인 URL
<https://rgai4phalo.labeling.us-east-1.sagemaker.aws> []

프라이빗 팀 (1) 정보
프라이빗 작업 인력의 작업 팀입니다. 레이블 지정 작업을 위해 여러 프라이빗 팀을 만들 수 있습니다.

작업자	삭제	프라이빗 팀 만들기			
이름	▼	ARN	▼	생성 시간	▼
private-team-1	arn:aws:sagemaker:us-east-1:556463194331:workteam/private-crowd/private-team-1			2020년 7월 29일 오전 8:57 UTC	

작업자 정보
프라이빗 작업 인력의 모든 작업자

작업자	활성화	비활성화	삭제	새 작업자 초대					
작업자 검색	< 1 ... >	[]	[]	[]					
이메일	▲	상태	▼	Cognito 상태	▼	활성	▼	사용자 이름	▼
ggamini@hmail.hanyang.ac.kr		초대장 보냄		Force_change_password		예		ggamini@hmail.hanyang.ac.kr	

<https://console.aws.amazon.com/sagemaker/groundtruth?region=us-east-1#>

© 2008 - 2020, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

- private team을 생성하면 다음과 같이 나옵니다.
- 메일로 들어가서 임시 password로 authorize를 해 줍니다.

The screenshot shows the AWS SageMaker Ground Truth interface. On the left, there's a sidebar with various options like 'Ground Truth', 'Notes', 'Training', etc. The main area displays information about a 'Private Team'. It shows the team name 'private-team-1', its ARN, and creation date ('2020년 7월 29일 오전 8:57 UTC'). Below this, there's a section for 'Workers' showing one worker entry with an email address 'ggamini@hmail.hanyang.ac.kr' and a status of 'Confirmed'.

- 위와 같이 authorized로 바뀝니다.

2. Create a labeling job

이제 labeling을 해보겠습니다.

The screenshot shows the 'Create Labeling Job' page in the AWS SageMaker console. The top navigation bar includes 'Amazon SageMaker Studio' and '레이블 지정 작업'. The main area has a search bar 'Search labeling jobs' and a button '레이블 지정 작업 생성'. A message at the bottom states '현재 생성된 레이블 지정 작업이 없습니다.' (No labeling jobs have been created yet).

- 레이블 지정 작업 생성으로 들어갑니다.

The screenshot shows the '작업 세부 정보 지정' (Job Sub-Information Configuration) step in the SageMaker labeling job creation process. The left sidebar shows 'Step 1: 작업 세부 정보 지정' and 'Step 2: 작업자와 구성 도구 선택'. The main area is titled '작업 개요' (Job Overview) and contains the following fields:

- 작업 이름:** job-demo-1
- I want to specify a label attribute name different from the labeling job name.** (checkbox)
- Input data setup:** Automated data setup (selected). Description: Provide the S3 location of the dataset you want labeled and let Ground Truth automatically connect to and use this dataset for your job.
- S3 location for input datasets:** s3://input-dataset-bucket/optional-prefix
- S3 location for output datasets:** Same location as input dataset (selected).
- Data type:** (dropdown menu)

At the bottom, there are links for '의견' (Feedback), '한국어' (Korean), and copyright information: © 2008 - 2020, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관.

- s3에 있는 파일로 할 수도, manifest file을 만들 수도 있습니다.

The screenshot shows the '버킷 만들기' (Create Bucket) wizard, step 1: 이름 및 지역 (Bucket Name and Region). The steps are numbered 1 through 4 at the top: 1. 이름 및 지역, 2. 옵션 구성, 3. 권한 설정, 4. 검토.

Fields in step 1:

- 버킷 이름:** hamin-groundtruth-test0-us
- 리전:** 미국 동부(버지니아 북부)
- 기존 버킷에서 설정 복사:** 버킷을 선택합니다(선택 사항). 4버킷

Buttons at the bottom: 생성 (Create), 취소 (Cancel), and 다음 (Next).

- S3는 다음과 같은 세팅을 하였습니다.

버킷 만들기

×

① 이름 및 지역 ② 옵션 구성 ③ 권한 설정 ④ 검토

속성

버전 관리

동일 버킷 내에 한 객체의 모든 버전을 보관합니다. [세부 정보 ↗](#)

서버 액세스 로깅

버킷에 대한 액세스 요청을 기록합니다. [세부 정보 ↗](#)

원본 버킷과 대상 버킷이 동일하면 버킷에 작성되는 로그에 대해 추가 로그가 생성됩니다. 이러한 추가 로그는 스토리지 결제 요금을 증가시키고 원하는 로그를 찾기가 힘들어질 수 있습니다.

대상 버킷

hamin-groundtruth-test0-us

대상 접두사

대상 접두사를 입력합니다.

태그

태깅을 사용하여 프로젝트 비용을 추적할 수 있습니다. [세부 정보 ↗](#)

키	값
+ 다른 항목 추가	

[이전](#) [다음](#)

- 해당 버킷에 대한 액세스 요청을 모두 기록합니다.

버킷 만들기

×

① 이름 및 지역 ② 옵션 구성 ③ 권한 설정 ④ 검토

참고: 버킷 생성 후에는 특정 사용자에게 액세스 권한을 부여할 수 있습니다.

퍼블릭 액세스 차단(버킷 설정)

모든 퍼블릭 액세스 차단

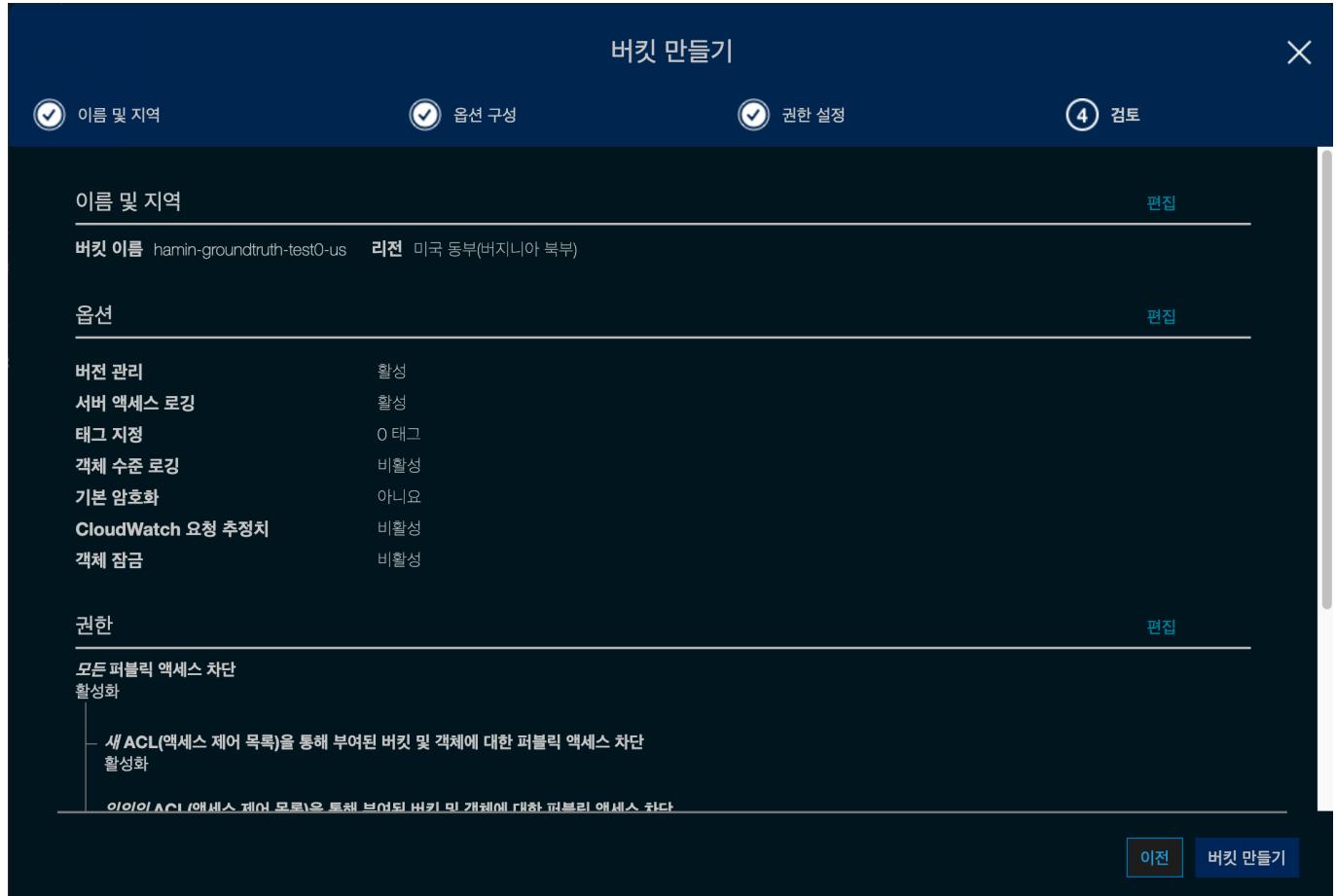
이 설정을 활성화하면 아래 4개의 설정을 모두 활성화한 것과 같습니다. 다음 설정 각각은 서로 독립적입니다.

- 새 ACL(액세스 제어 목록)을 통해 부여된 버킷 및 객체에 대한 퍼블릭 액세스 차단
S3은 새로 추가된 버킷 또는 객체에 적용되는 퍼블릭 액세스 권한을 차단하며, 기존 버킷 및 객체에 대한 새 퍼블릭 액세스 ACL 생성을 금지합니다. 이 설정은 ACL을 사용하여 S3 리소스에 대한 퍼블릭 액세스를 허용하는 기존 권한을 변경하지 않습니다.
- 임의의 ACL(액세스 제어 목록)을 통해 부여된 버킷 및 객체에 대한 퍼블릭 액세스 차단
S3은 버킷 및 객체에 대한 퍼블릭 액세스를 부여하는 모든 ACL을 무시합니다.
- 새 퍼블릭 버킷 또는 액세스 지점 정책을 통해 부여된 버킷 및 객체에 대한 퍼블릭 액세스 차단
S3는 버킷 및 객체에 대한 퍼블릭 액세스를 부여하는 새 버킷 및 액세스 지점 정책을 차단합니다. 이 설정은 S3 리소스에 대한 퍼블릭 액세스를 허용하는 기존 정책을 변경하지 않습니다.
- 임의의 퍼블릭 버킷 또는 액세스 지점 정책을 통해 부여된 버킷 및 객체에 대한 퍼블릭 및 교차 계정 액세스 차단
S3는 버킷 및 객체에 대한 퍼블릭 액세스를 부여하는 정책을 사용하는 버킷 또는 액세스 지점에 대한 퍼블릭 및 교차 계정 액세스를 무시합니다.

시스템 권한 관리

[이전](#) [다음](#)

- 권한 설정도 우선 기본 사항으로 하였습니다.



- 이렇게 S3 버킷을 생성했습니다.
- 그리고 샘플 dataset은 제 마음대로 넣어봤습니다.
- S3 버킷에 파일을 넣고 manifest를 생성해주어야 합니다.
- manifest 만들기는 다음을 참고합니다

https://docs.aws.amazon.com/ko_kr/sagemaker/latest/dg/sms-data-input.html#sms-console-create-manifest-file

<https://docs.aws.amazon.com/sagemaker/latest/dg/sms-getting-started-step1.html>

- manifest는 샘플들을 리스트하는 json 파일인데 docs와 다르게 manifest 없이도 할 수 있도록 되어있습니다.

그리고 다시 labeling으로 돌아가

Amazon SageMaker Studio

Ground Truth

Input data setup

S3 location for input datasets

S3 location for output datasets

Data type

- 이런식으로 세팅을 해 줍니다

Select the type of data being labeled to view available task templates for it or select 'Custom' to create your own.

Image

작업 선택

Image 분류(다중 레이블)

경계 상자

의미 체계 분할

- dataset이 이미지를 이므로 task category는 image, task selection은 semantic segmentation으로 해보겠습니다.

다음으로 넘어가서

- private으로 세팅해 줍니다.

- 이런식으로 예시가 나옵니다
- 구분해야할 물체를 최대한 구체적으로 명시해줍니다.

Amazon SageMaker Studio

Amazon SageMaker > 레이블 지정 작업

작업	레이블 지정 작업
C	작업
Search labeling jobs	레이블 지정 작업 생성
이름 상태 작업 유형 job-demo-1 진행 중 의미 체계 분할 2020년 7월 30일 오전 9:19 UTC	

© 2008 - 2020, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

- 생성 버튼을 눌러주면 다음과 같이 labeling job이 생성 됩니다.

Amazon SageMaker Studio

Amazon SageMaker > 레이블 지정된 데이터 세트 객체 (-)

출력 | 태그

작업	레이블 지정된 데이터 세트 객체 (-)
pic1.jpg	
pic2.jpg	
pic3.jpg	
pic4.jpg	

© 2008 - 2020, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

- 들어가보면 작업 중인 사진들이 보입니다.
- 시간이 좀 걸리니 기다려야 합니다.

Amazon SageMaker Studio

레이블 지정 작업 정보

이름	상태	작업 유형	레이블이 지정된 객체/전체	생성 시간
job-demo-1	진행 중	의미 체계 분할	0 / 8	2020년 7월 30일 오전 9:19 UTC

© 2008 - 2020, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

- 결과물을 보려면 console로 들어가야 합니다.

Amazon SageMaker Studio

프라이빗 작업 인력 요약 정보

Amazon Cognito 사용자 풀
us-east-1_BOqZg77p8 []

엘 클라우언트
1d3272645r7ej94hr18uaedque []

프라이빗 팀 (1) 정보

작업자	생성 시간
private-team-1	2020년 7월 29일 오전 8:57 UTC

작업자 정보

작업자 검색

이메일 상태 Cognito 상태 활성 사용자 이름

© 2008 - 2020, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

- 오른쪽의 레이블 지정 포털 로그인 URL로 들어가면 됩니다.

Amazon SageMaker Ground Truth - Worker Portal

Hello, ggaminig@mail.hanyang.ac.kr

Log out Show instructions Start working

Jobs

Description	Customer ID	Creation time
Semantic segmentation: please segment people and traffic light and bus, tree, car	556463194331	July 30, 2020 9:21:21 UTC

- 들어가면 다음과 같은 화면이 나옵니다.
- start working을 들어가면

Amazon SageMaker Ground Truth - Worker Portal

Hello, ggaminig@mail.hanyang.ac.kr Customer ID: 5564... Task description: Draw pixel level labels around specific objects and... Task time: 0:42 of 60 Min Stop working Log out

Instructions: please segment people and traffic light and bus, tree, car

Instructions View full instructions View tool guide How to use the Auto-segment tool Good example Enter description to explain a correctly done segmentation Add image here Bad example Enter description of an incorrectly done segmentation Add image here

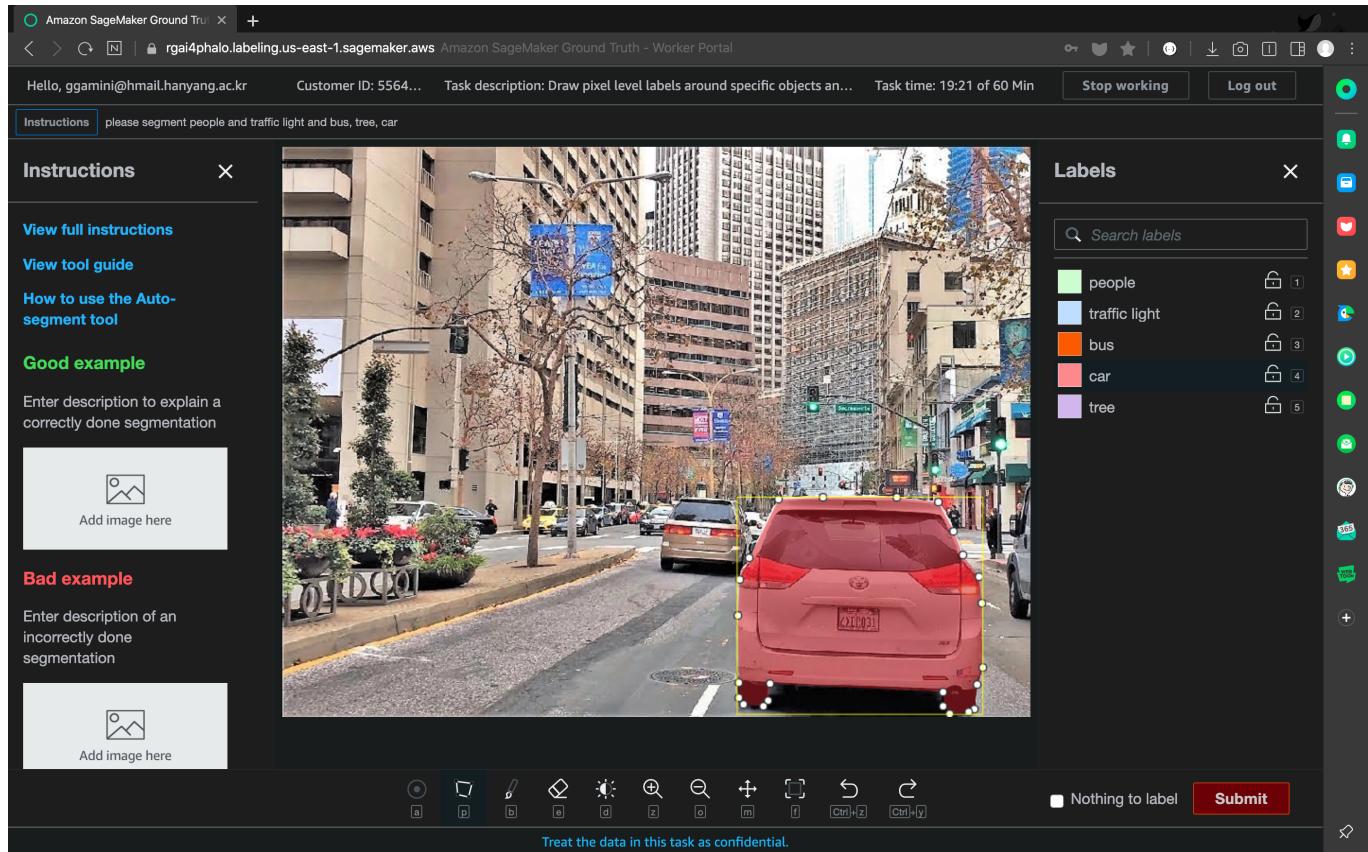
Labels Search labels

- people
- traffic light
- bus
- car
- tree

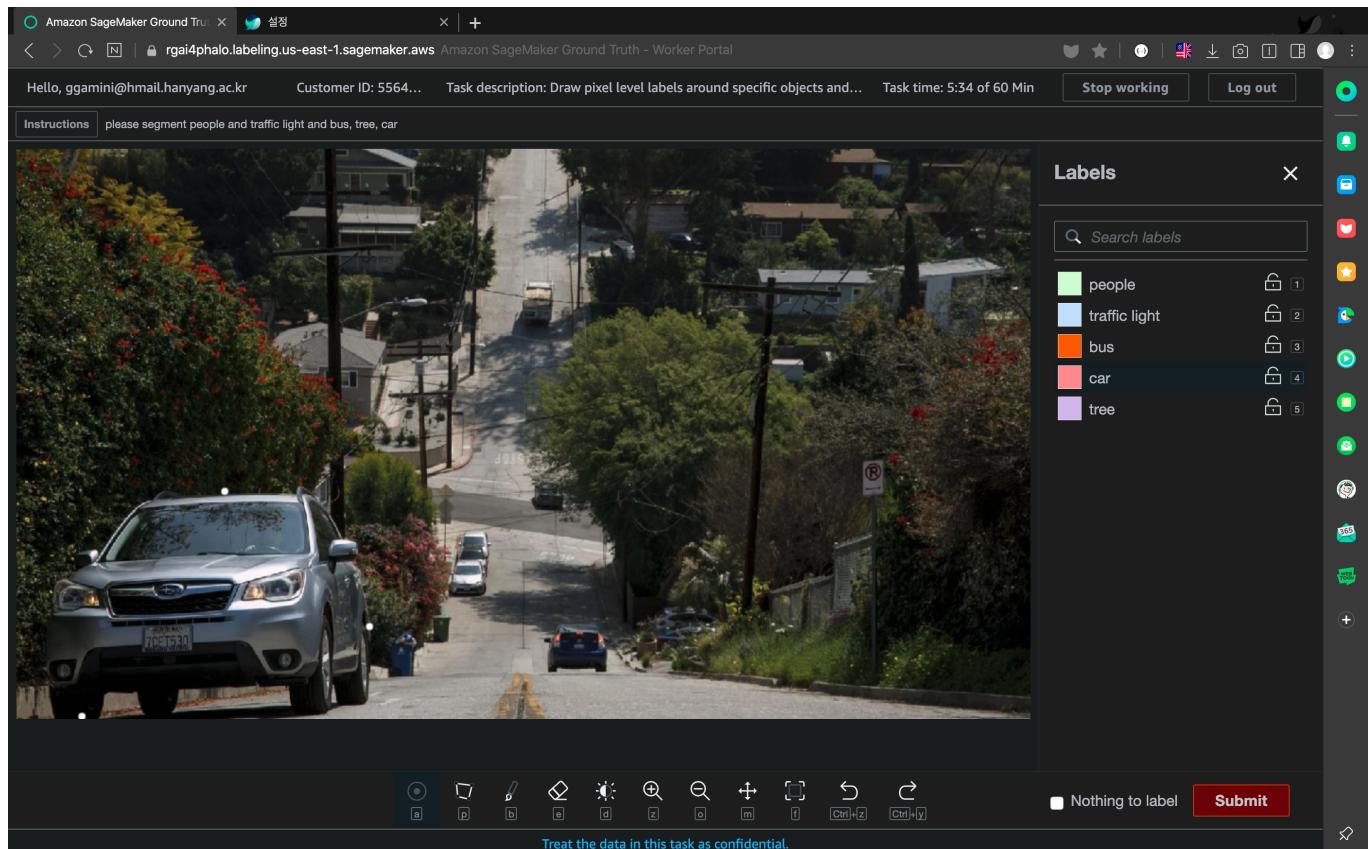
Submit Nothing to label

Treat the data in this task as confidential.

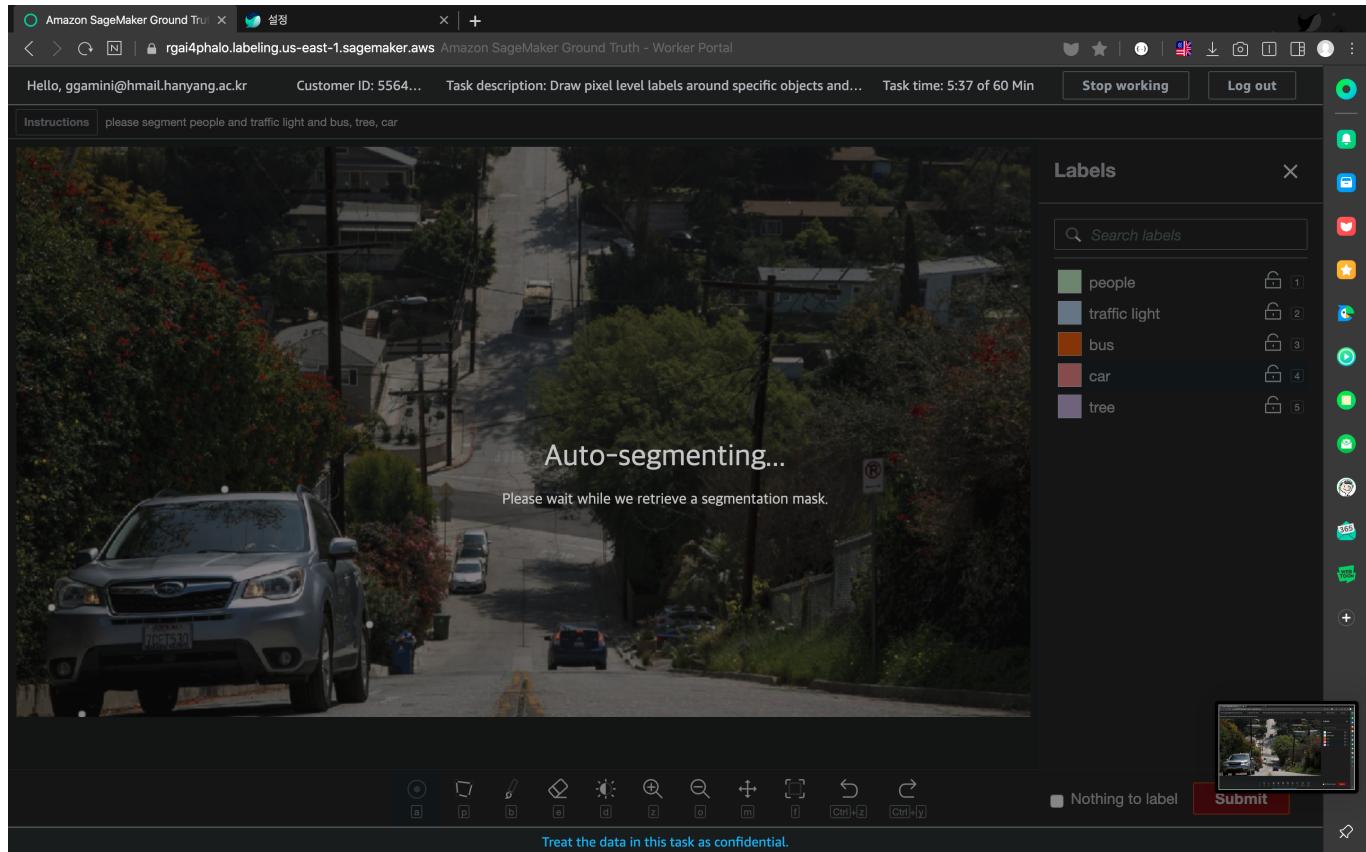
- 다음과 같은 화면이 나옵니다.



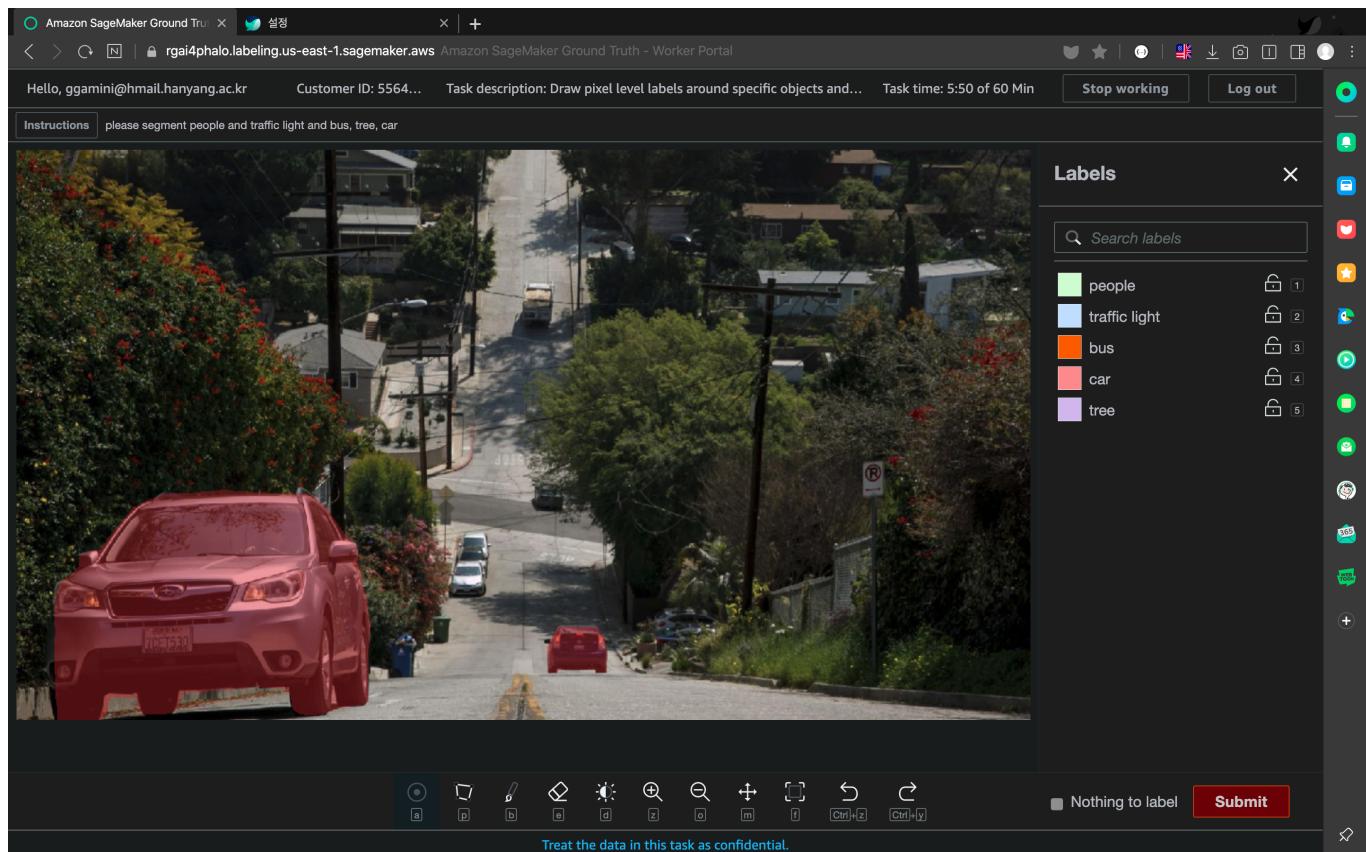
- 이런식으로 지정해줍니다
- 그리고 submit을 해줍니다.



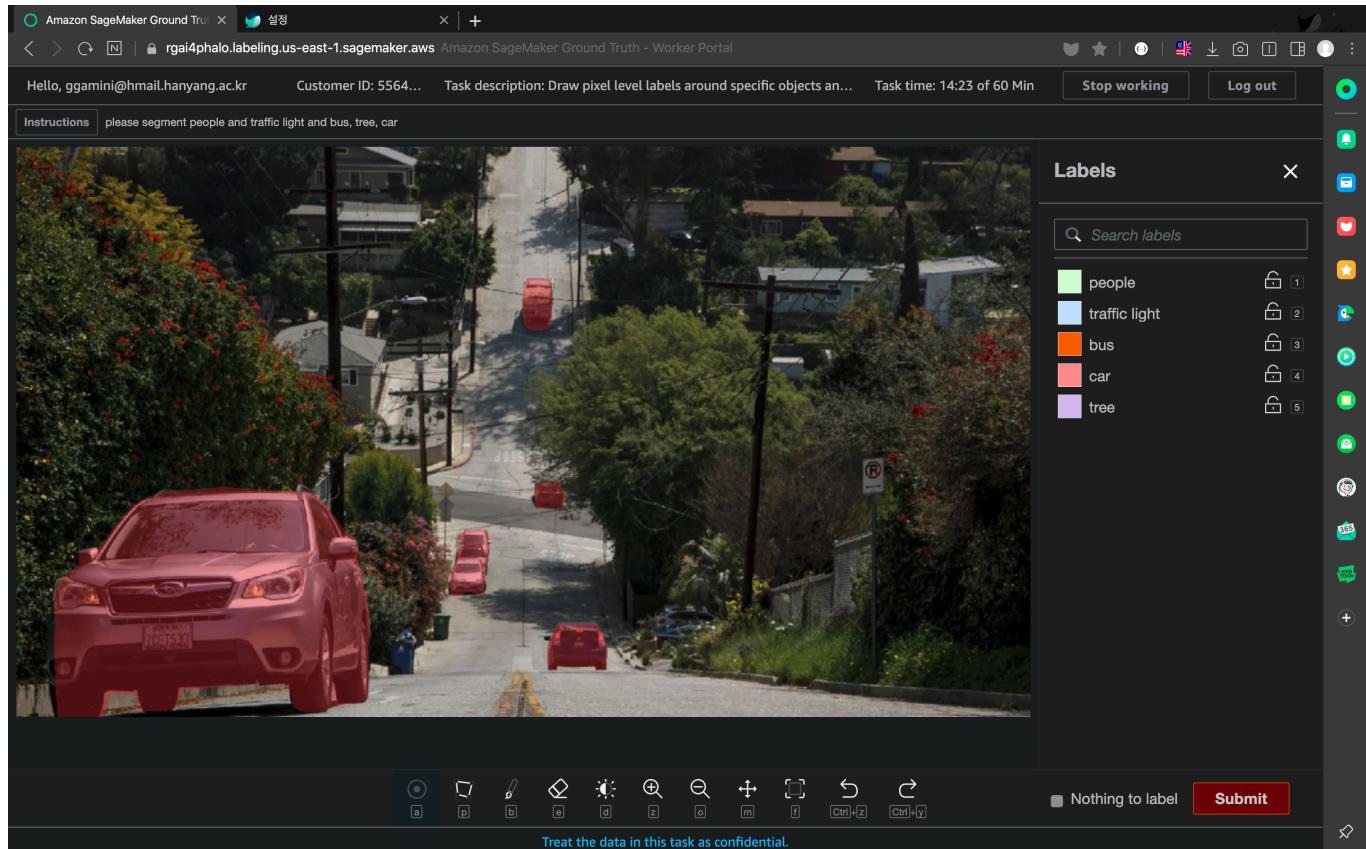
- 이번엔 auto segment 기능을 이용해보겠습니다.



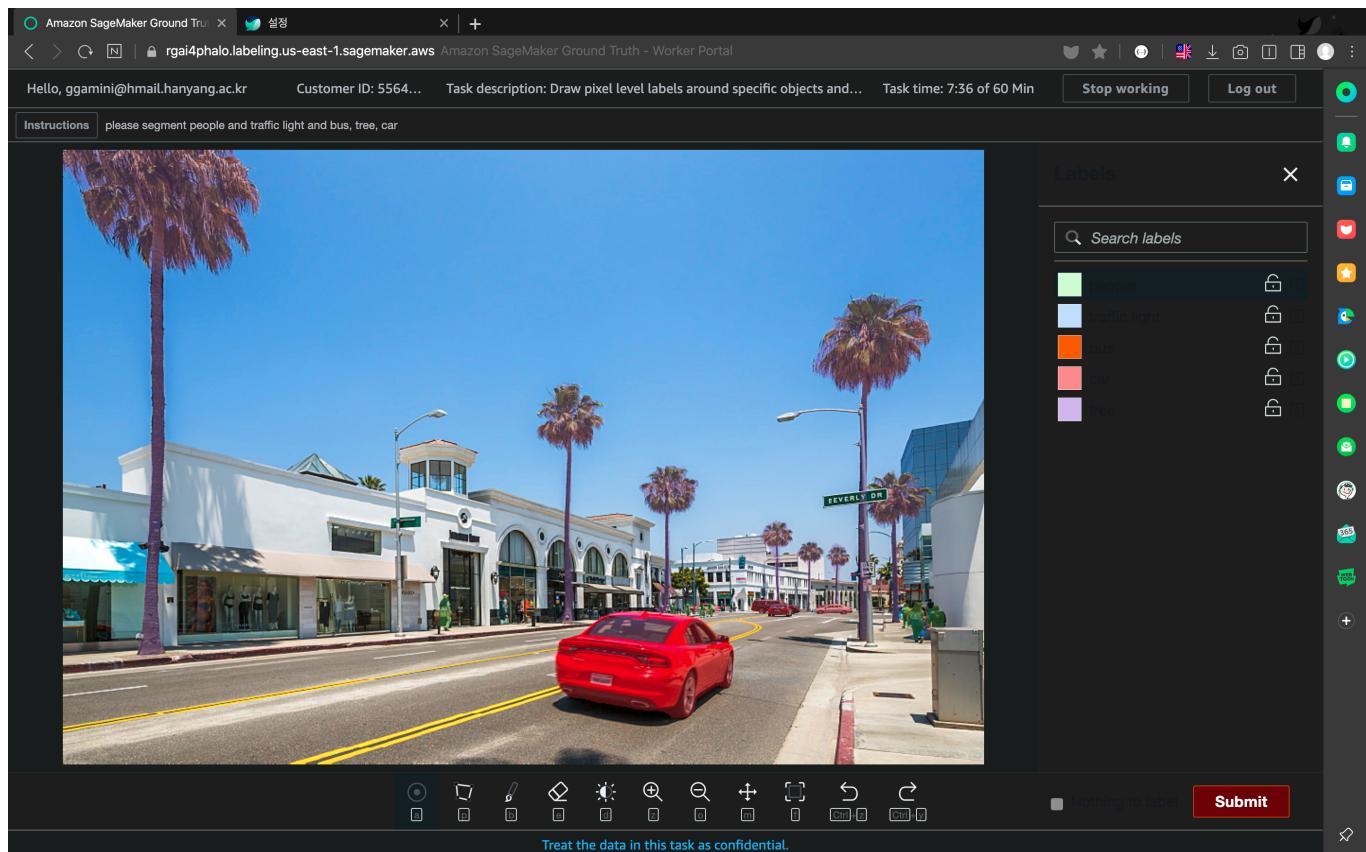
- Labels에서 해당되는것을 클릭후 최하단, 최상단, 좌우 끝 네 지점을 지정해주면 auto segmenting이 실행됩니다.



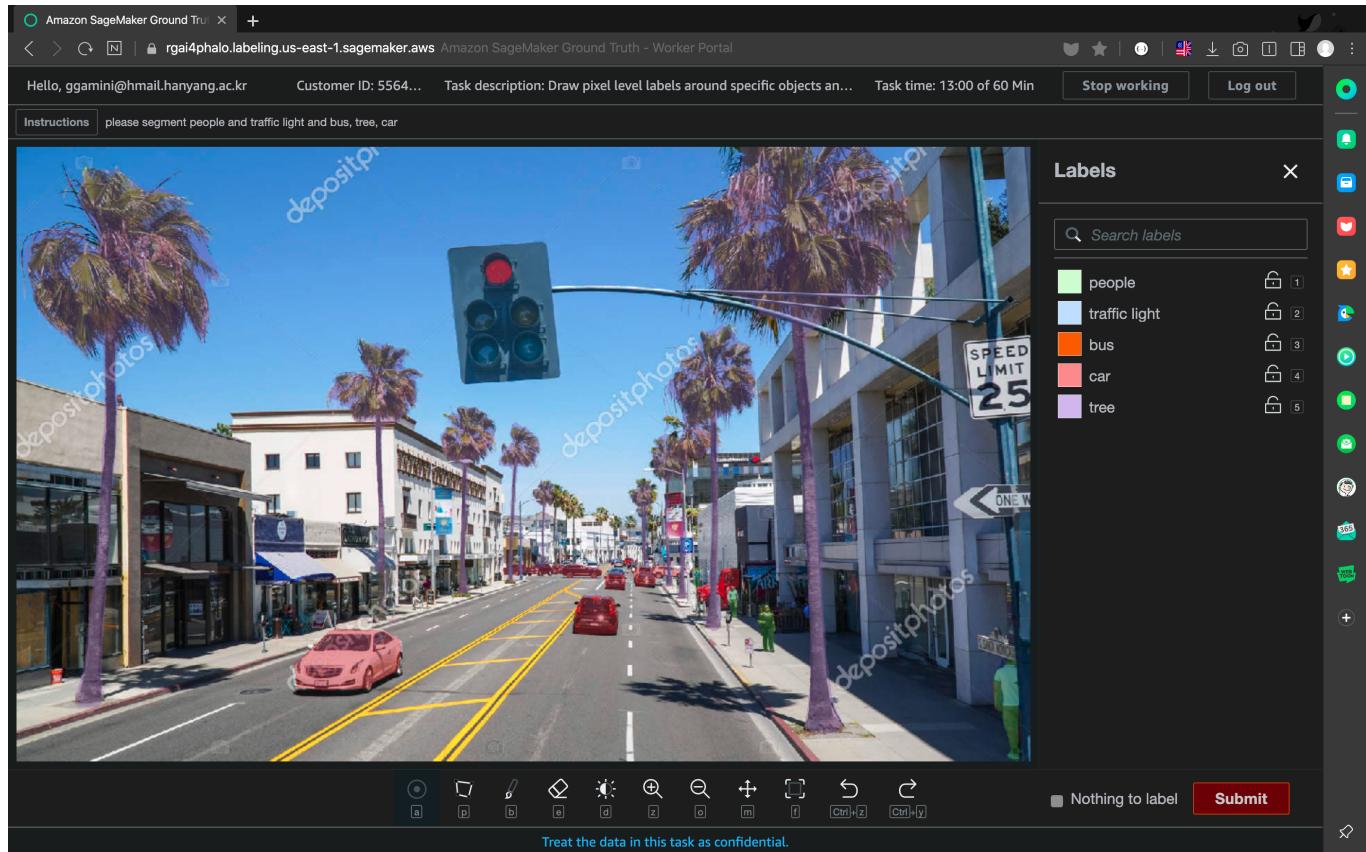
- 이런식으로 잘 지정이 됩니다.



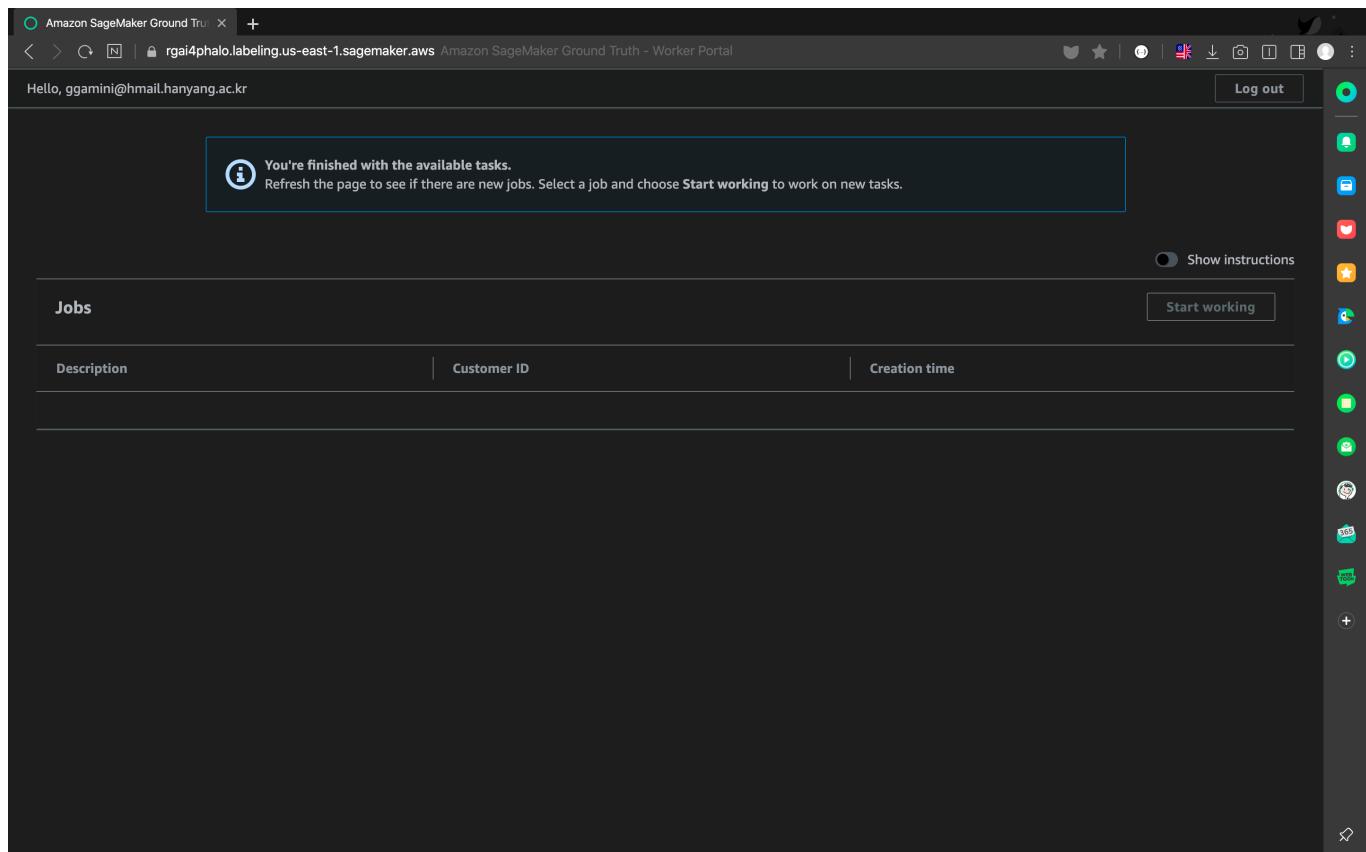
- 다른 차들도 해줍니다.



- 다른 요소들도 해줍니다.



- 다 해줍니다.



- 모두 마무리하면 이런 화면으로 넘어갑니다.

The screenshot shows the AWS CloudWatch Management Console interface. On the left, there's a sidebar with various AWS services like CloudWatch, ServiceLens, Container Insights, and Synthetics. The main area is titled 'CloudWatch Logs' and shows a log group for '/aws/sagemaker/LabelingJobs'. A specific log stream named 'job-demo-1' is selected. The interface displays a list of log events, each with a timestamp and a detailed log message. One event is expanded to show its full JSON structure. At the bottom right of the main area, there's a note about using CloudWatch Logs Insights.

- cloud watch에 들어가면 실행한 log를 볼 수 있습니다.

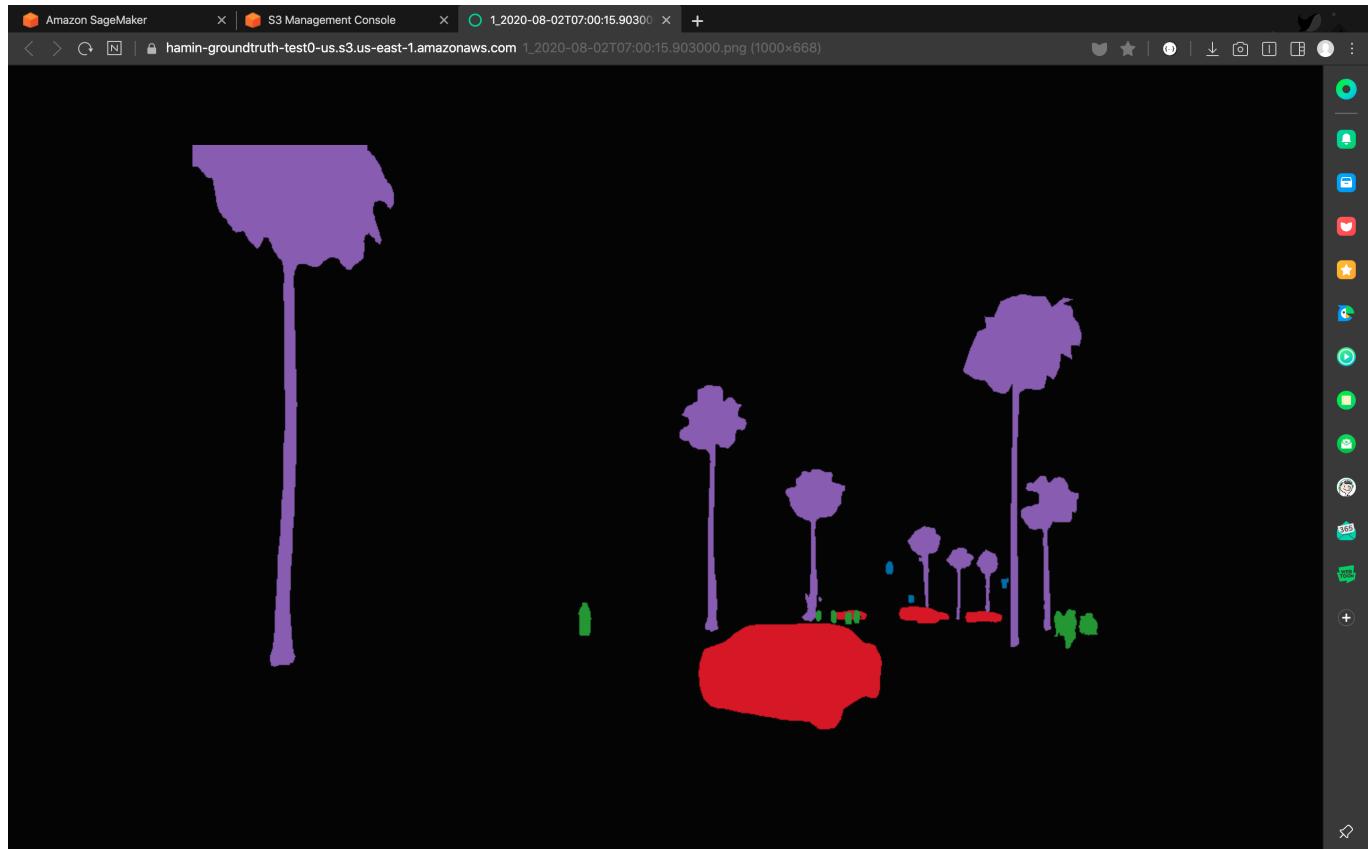
The screenshot shows the AWS S3 Management Console interface. It displays a bucket named 'hamin-groundtruth-test0-us'. Under the 'output' folder, several files are listed, including 'index', 'input', and several timestamped files like '2020-08-02-07-40-06-A11479F487ADA293'. Each file entry shows its last modified date, size, and storage class (Standard). The top navigation bar includes tabs for '개요', '속성', '권한', '관리', and '액세스 지점'.

- S3의 output에 들어가면 결과물이 들어있습니다.

Screenshot of the AWS S3 Management Console showing a list of annotated images. The list includes 8 files, all of which are standard storage classes.

이름	마지막 수정	크기	스토리지 클래스
0_2020-08-02T07:38:26.700000.png	8월 2, 2020 4:38:27 오후 GMT+0900	12.8 KB	стандарт
1_2020-08-02T07:00:15.903000.png	8월 2, 2020 4:00:16 오후 GMT+0900	4.8 KB	стандарт
2_2020-08-02T07:14:42.914000.png	8월 2, 2020 4:14:43 오후 GMT+0900	11.3 KB	стандарт
3_2020-08-02T06:30:15.279000.png	8월 2, 2020 3:30:16 오후 GMT+0900	937.0 B	стандарт
4_2020-08-02T07:31:14.293000.png	8월 2, 2020 4:31:15 오후 GMT+0900	31.6 KB	стандарт
5_2020-08-02T06:45:46.945000.png	8월 2, 2020 3:45:47 오후 GMT+0900	2.0 KB	стандарт
6_2020-08-02T07:17:49.378000.png	8월 2, 2020 4:17:50 오후 GMT+0900	3.4 KB	стандарт
7_2020-08-02T06:51:59.428000.png	8월 2, 2020 3:52:00 오후 GMT+0900	3.0 KB	стандарт

- 결과물들은 이곳에 있습니다.



- 이런식으로 있습니다.

The screenshot shows the Amazon SageMaker Studio interface. On the left, a sidebar lists various project categories: Ground Truth (Labeling Job, Dataset with Labeling, Labeling Inference), Notebooks (Notebook Instances, Configuration, Git Repository), Processing (Job Processing, Model Packaging, Model), and Training (Training Job, Model Training, Model Tuning). The main content area displays a 'job-demo-1' job. It includes sections for 'Labeling Job Details' (Input Dataset Location: s3://hamin-groundtruth-test0-us/input/index/dataset.manifest.json, Output Dataset Location: s3://hamin-groundtruth-test0-us/output/job-demo-1/), 'ARN' (arn:aws:sagemaker:us-east-1:55646319431:labeling-job/job-demo-1), and 'Job Type' (Private). Below this, a 'Labeling Data Set Objects (8)' section shows four preview images: a city street with green objects, a red car on a street, a city street with palm trees, and a red car from behind.

- 이곳에 가면 output의 경로를 볼 수 있습니다.

The screenshot shows the Amazon SageMaker Studio interface, similar to the previous one but focusing on labeled dataset objects. The main content area displays a 'Labeled Data Set Objects (8)' section. It shows four images with their corresponding labels: 'pic1.jpg' (a city street with green objects), 'pic2.jpg' (a red car on a street), 'pic3.jpg' (a city street with palm trees), and 'pic4.jpg' (a red car from behind). The labels are displayed as semi-transparent purple overlays on the images.

- 이런식으로 잘 labeling이 되어있습니다.