

본 실습에서는 초음파 사진을 통해 유방암을 진단하는 모델을 Backend.AI를 통해 학습시키는 방법을 다룹니다.

데이터셋 다운로드

 ARYA SHAH · UPDATED 2 YEARS AGO

▲ 184

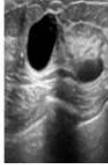
New Notebook

 Download (204 MB)



Breast Ultrasound Images Dataset

Breast ultrasound images for classification, detection & segmentation



Data Card

Code (75)

Discussion (2)

About Dataset

Breast cancer is one of the most common causes of death among women worldwide. Early detection helps in reducing the number of early deaths. The data reviews the medical images of breast cancer using ultrasound scan. Breast Ultrasound Dataset is categorized into three classes: normal, benign, and malignant images. Breast ultrasound images can produce great results in classification, detection, and segmentation of breast cancer when combined with machine learning.

Data

The data collected at baseline include breast ultrasound images among women in ages between 25 and 75 years old. This data was collected in 2018. The number of patients is 600 female patients. The dataset consists of 780 images with an average image size of 500*500 pixels. The images are in PNG format. The ground truth images are presented with original images. The images are categorized into three classes, which are normal, benign, and malignant.

If you use this dataset, please cite:

Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863.

Usability ⓘ

10.00

License

CC0: Public Domain

Expected update frequency

Never

<https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset>에 접속하여 다운로드 버튼을 클릭해 데이터셋을 다운로드 받습니다. Kaggle에 로그인되어있지 않다면 로그인을 해야 합니다.



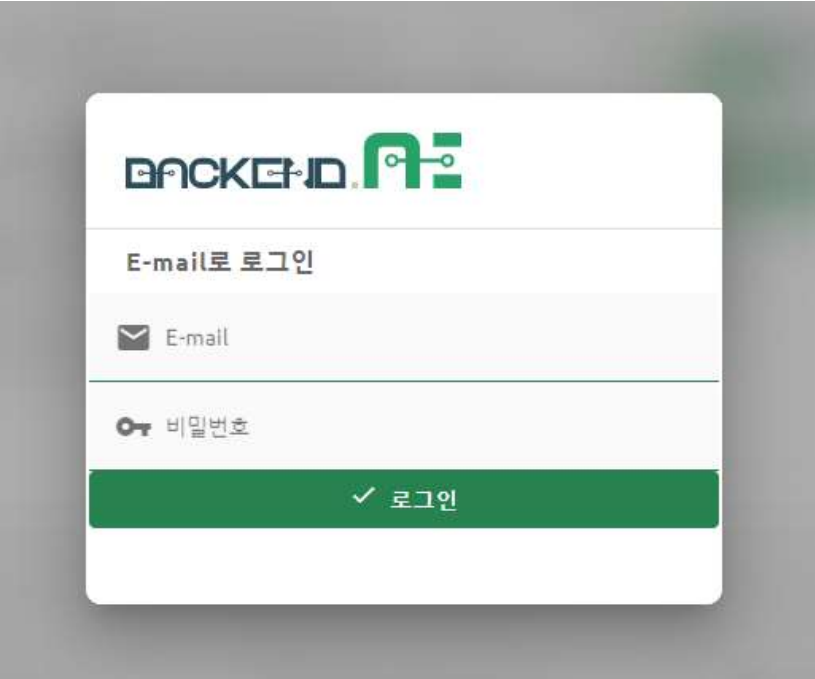
› Users › skku › 다운로드 › archive ›

이름

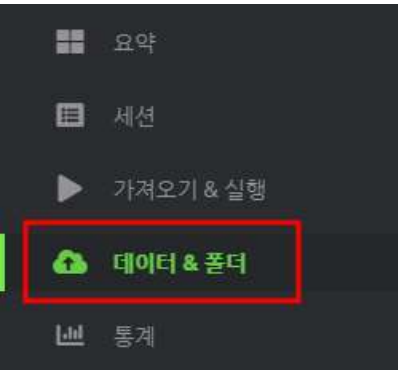
 Dataset_BUSI_with_GT

다운로드 받은 파일을 압축해제하면 폴더 내에 Dataset_BUSI_with_GT 라는 폴더가 있는 것을 확인할 수 있습니다.

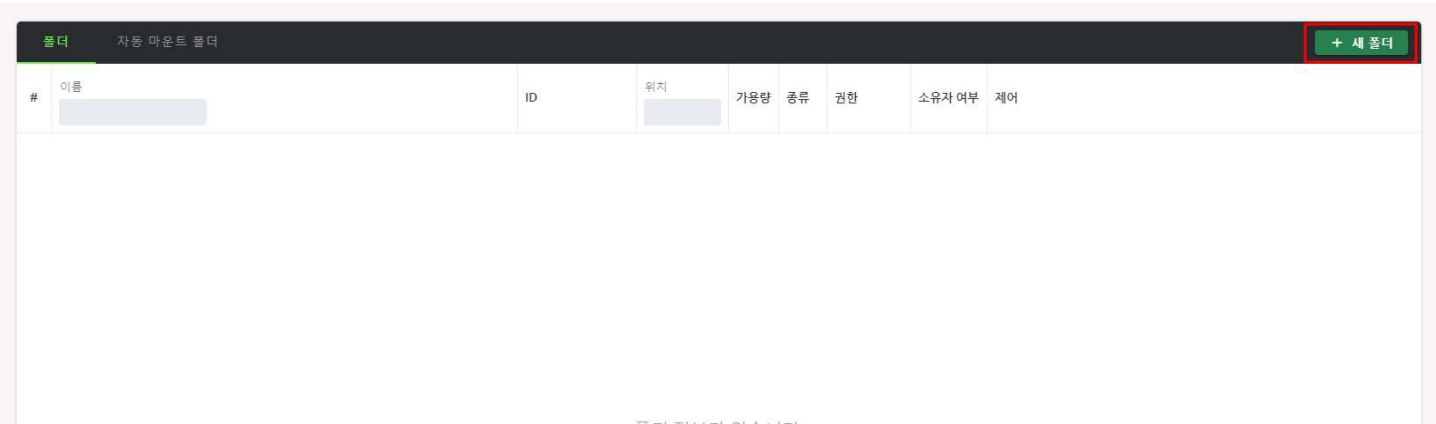
데이터셋 импорт



웹 브라우저를 통해 <https://spckcluster.skku.edu/>에 접속하여 로그인합니다. (본 실습에서는 데스크탑 버전이 아닌 웹 버전을 사용합니다.)



데이터 & 폴더 탭을 클릭합니다.



새 폴더 버튼을 클릭하여 새로운 폴더를 생성합니다.

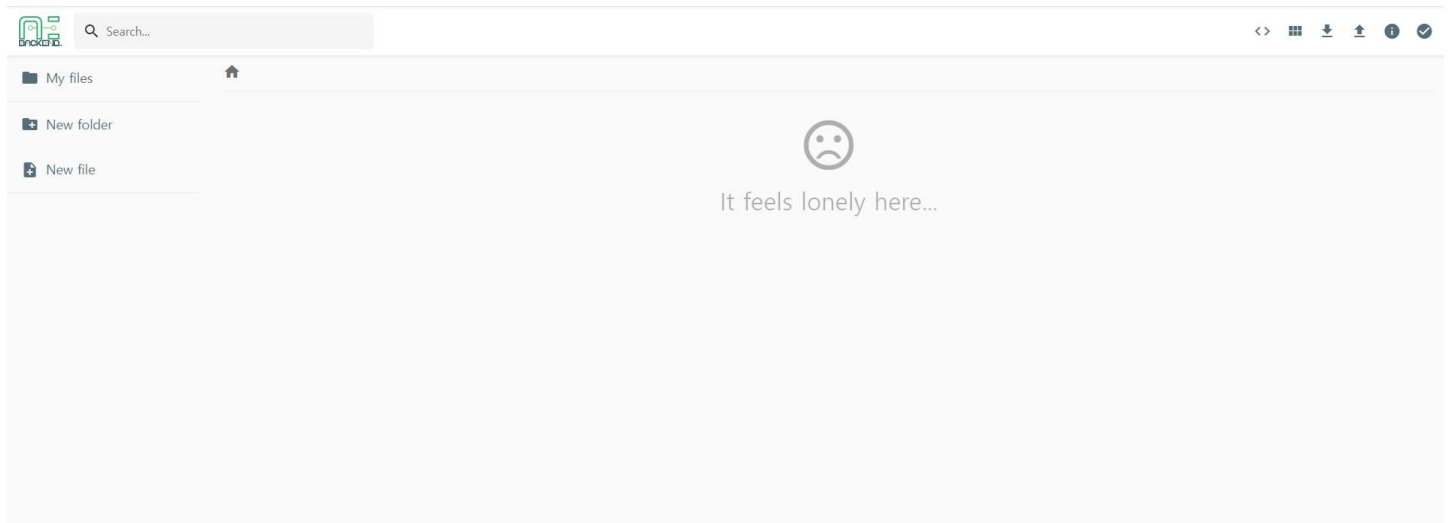
는 것이 편리합니다. **파일브라우저 실행** 버튼을 클릭하여 파일 브라우저를 실행합니다.

i 파일 브라우저

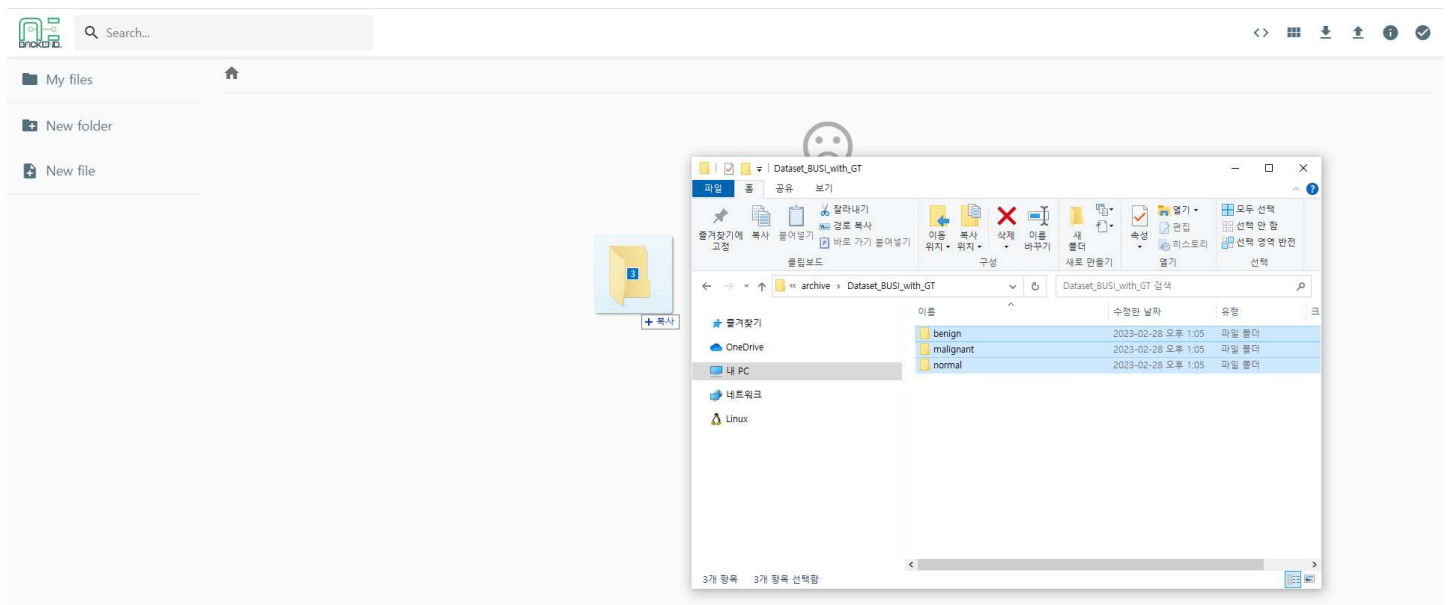
파일 브라우저는 웹 브라우저를 통해 원격 서버에 있는 파일 관리를 도와주는 프로그램입니다. 특히 사용자의 로컬 머신에 있는 디렉토리를 업로드 할 때에도 유용하게 사용할 수 있습니다.

현재 Backend.AI에서는 파일 브라우저를 연산 세션 내에서 실행되는 어플리케이션 형태로 제공합니다. 따라서, 다음과 같은 최소 조건이 필요합니다.

- 최소 1개 이상의 세션 생성 가능
- 최소 CPU 1 core, RAM 0.5 GB 이상의 여유 자원



파일 브라우저를 실행하면 웹 브라우저의 새 탭이 열리면서 사진과 같은 화면이 나타납니다.



압축 해제한 폴더 내에 있는 세 개의 폴더를 선택한 후 드래그하여 파일 브라우저 화면에 놓아 업로드합니다. 업로드에는 약간의 시간이 소요될 수 있습니다.

#	이름	생성일	크기	작업
1	benign	Tue, 28 Feb 2023 13:37:26 GMT	65536	
2	malignant	Tue, 28 Feb 2023 13:37:29 GMT	32768	
3	normal	Tue, 28 Feb 2023 13:37:29 GMT	16384	

다시 Backend.AI으로 돌아와 폴더를 열어보면 업로드한 폴더가 있는 것을 확인할 수 있습니다.

모델 импорт

학습 모델은 https://github.com/neowizard2018/neowizard/blob/master/DeepLearningProject/Project_LEC_01_Example.ipynb를
임포트하여 사용할 것입니다.

neowizard2018 / neowizard Public

<> Code Issues Pull requests Actions Projects Security Insights

master 1 branch 0 tags

neowizard2018 1st commit

- DataAnalysis 1st commit
- DeepLearningProject 1st commit
- MachineLearning hidden nodes 8 => 30
- TensorFlow2 2nd commit
- README.md Update README.md

README.md

YouTube 강의 사이트 : <https://www.youtube.com/neowizard>

Clone

HTTPS SSH GitHub CLI

<https://github.com/neowizard2018/neowizard.git>

Use Git or checkout with SVN using the web URL.

Open with GitHub Desktop

Open with Visual Studio

Download ZIP

<https://github.com/neowizard2018/neowizard>에 접속하여 Code 버튼을 클릭하여 GitHub 주소를 복사합니다.



다시 Backend.ai에 접속하여 가져오기 & 실행 탭을 클릭합니다.

GitHub 저장소 가져오기

가져온 GitHub 저장소는 폴더로 변환되며, 세션을 시작할 때 마운트할 수 있습니다.

GitHub 주소

https://github.com/neowizard2018/neowizard.git

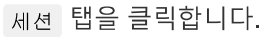
호스트

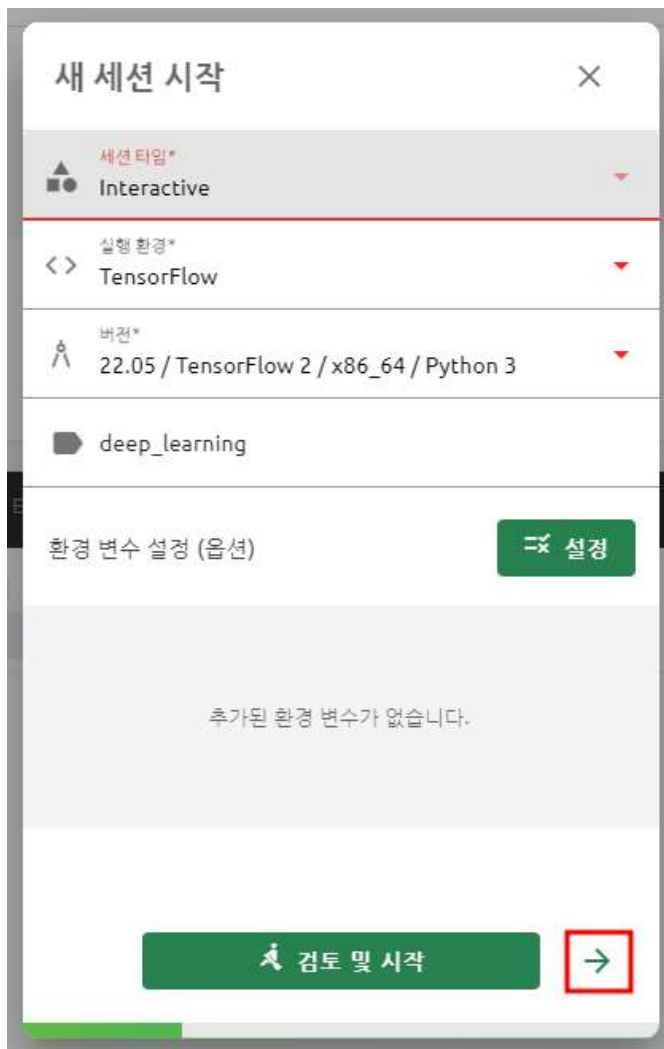
sp-bai-sc-g010:sas

가져오기

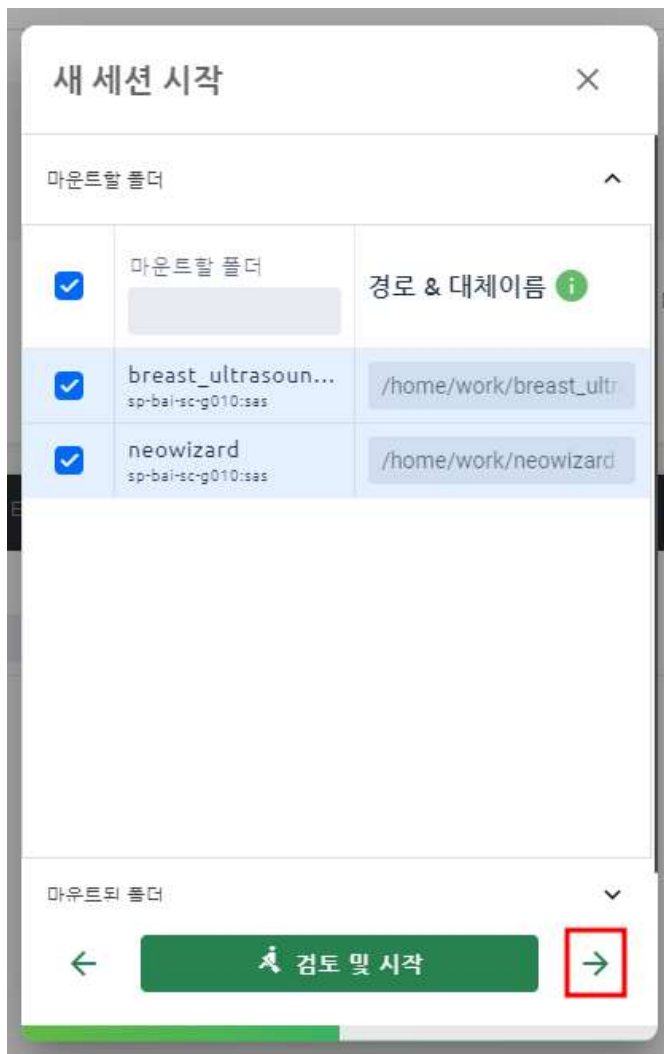
폴더 자동 마운트 폴더								+ 새 폴더
#	이름	ID	위치	가용량	종류	권한	소유자 여부	제어
1	 breast_ultrasound_dataset	8acf780375cc40da 8dd442bd278ff64b9	sp-bal-sc- g010:sas	-		 		     
2	 newwizard	a07b6a5d01cae4b08 99013a314fd64fdaf	sp-bal-sc- g010:sas	-		 		     

세션 생성





새로운 세션을 생성할 수 있는 대화 상자가 나타나는 것을 확인할 수 있습니다. 세션 타입은 `Interactive`, 실행 환경은 `TensorFlow`, 세션 이름은 `deep_learning` 으로 설정하겠습니다. 화살표 버튼을 클릭하여 다음 단계로 진행합니다.



다음 단계에서는 마운트할 폴더를 설정할 수 있습니다. 이전에 업로드한 `breast_ultrasound_dataset` 폴더와 `neowizard` 폴더를 마운트합니다. 또한, 마운트되는 상위 폴더의 기본 경로는 `/home/work` 인 것을 확인할 수 있습니다. 화살표 버튼을 클릭하여 다음 단계로 진행합니다.

새 세션 시작

×

자원 그룹*

default

▼

자원 할당*

large-gpu (4CPU 8GB 64MB 4 GPU)

▼

사용자 설정 자원 할당

▼

클러스터 모드 설정*

단일 노드

▼

클러스터 크기

1 컨테이너

고성능 컴퓨팅 최적화

▼

←

검토 및 시작

→

다음 단계에선 세션에 할당할 자원을 설정할 수 있습니다. 본 실습에서는 `large-gpu (4CPU 8GB 64MB 4 GPU)` 를 선택하겠습니다. 원한다면, 사용자 자원 할당을 통해 커스텀 자원을 설정할 수도 있습니다. 클러스터 모드는 `단일 노드` 로 설정하겠습니다. 화살표 버튼을 클릭하여 마지막 단계로 진행합니다.

새 세션 시작

×

세션 정보

환경 정보

TENSORFLOW (NGC) 22.05

TENSORFLOW 2X86_64PYTHON 3

INTERACTIVE

프로젝트 이름Edu

자원 그룹default

총 자원 할당량

CPU	메모리	공유메모리	GPU
4	8.0	64.00	4
Core	GB	MB	슬롯

×1

컨테이너

단일

노드

배치

마운트된 폴더

breast_ultrasound_dataset (→ /home/work/breas...

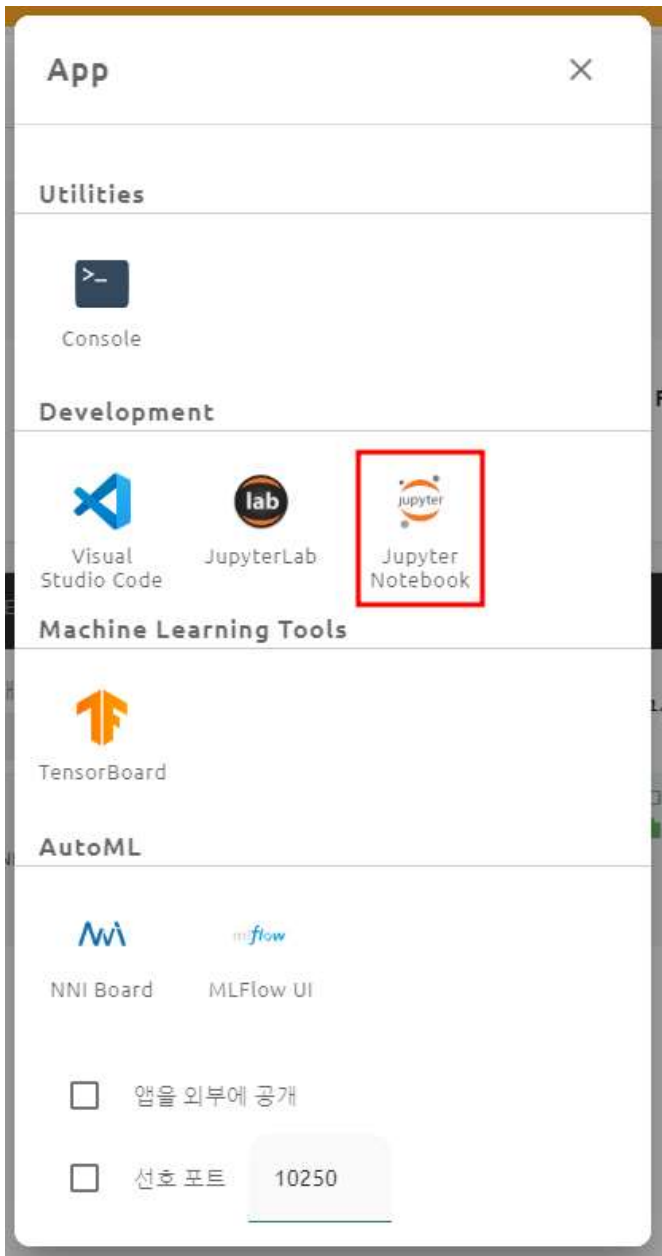
neowizard (→ /home/work/neowizard)

.test (Automount folder)

←

시작

마지막 단계에서는 지금까지 설정한 세션의 요약 정보를 확인할 수 있습니다. 뒤로 가기 버튼을 클릭하여 설정을 수정할 수도 있습니다, 설정을 마쳤다면 시작 버튼을 클릭하여 세션을 생성합니다.



세션 생성이 완료되면 세션을 실행하기 위한 App을 선택할 수 있는 대화 상자가 나타납니다. Jupyter Notebook 을 선택합니다.

i **세션 시작**

실행중
INTERACTIVE
BATCH
종료
기타

시작

#	세션 정보	상태	제어	구성	사용량	예약시간	아키텍처	세션 타입
1	deep_learning TensorFlow 22.05 Nvidia GPU Cloud	RUNNING		breast_ultrasound default 4코어 8.00GB 4.00CPU	CPU RAM CPU I/O 0.0MB / W: 0.2	2023. 3. 3. 오전 10:06:47 (00:00:12)	x86_64	INTERACTIVE

세션을 나가더라도 실행중 탭에서 Apps 아이콘을 클릭하여 다시 세션을 시작할 수 있습니다.

Files Running IPython Clusters Nbextensions

Select items to perform actions on them. UPLOAD NEW ↺

	NAME	LAST MODIFIED	FILE SIZE
☐	0	/	
☐	breast_ultrasound_dataset	2일 전	
☐	neowizard	21분 전	
☐	DO_NOT_STORE_PERSISTENT_FILES_HERE.md	2달 전	540 B
☐	id_container	4분 전	1.68 kB

Jupyter Notebook이 실행되면 새 탭이 열리면서 사진과 같은 화면이 나타나는 것을 확인할 수 있습니다. 또한, 세션 생성시 마운트한 `breast_ultrasound_dataset` 폴더와 `neowizard` 폴더를 확인할 수 있습니다.

	NAME	LAST MODIFIED	FILE SIZE
☐	0	/ neowizard / neowizard-master / DeepLearningProject	
☐	..	몇 초 전	
☐	Project_LEC_01_Example.ipynb	Running 3달 전	374 kB
☐	Project_LEC_02_Example.ipynb	3달 전	405 kB
☐	readme.md	3달 전	1 B

/neowizard/neowizard-master/DeepLearningProject/Project_LEC_01_Example.ipynb 파일을 클릭하여 열어보겠습니다.

```
In [ ]: import tensorflow as tf
from tensorflow.keras.layers import Dropout, Dense, GlobalAveragePooling2D
from tensorflow.keras.applications import MobileNet
from tensorflow.keras.models import Sequential
from tensorflow.keras.preprocessing.image import ImageDataGenerator
```

Google Drive 마운트해서 데이터셋 가져옴

```
In [ ]: from google.colab import drive
drive.mount('/content/gdrive/')

Mounted at /content/gdrive/
```

ROOT_DIR, DATA_ROOT_DIR 설정

```
In [ ]: import os

ROOT_DIR = '/content'

DATA_ROOT_DIR = os.path.join(ROOT_DIR, 'Dataset_BUSI_with_GT')

CLASSIFICATION_DATA_ROOT_DIR = os.path.join(ROOT_DIR, 'Classification')

CLASSIFICATION_TRAIN_DATA_ROOT_DIR = os.path.join(CLASSIFICATION_DATA_ROOT_DIR, 'train')
```

```
In [ ]: import shutil

if os.path.exists(DATA_ROOT_DIR):
    shutil.rmtree(DATA_ROOT_DIR)

if os.path.exists(CLASSIFICATION_DATA_ROOT_DIR):
    shutil.rmtree(CLASSIFICATION_DATA_ROOT_DIR)
```

Google Drive 에서 dataset download

```
In [ ]: import shutil
```

```
import os

ROOT_DIR = '/home/work'

DATA_ROOT_DIR = os.path.join(ROOT_DIR, 'breast ultrasound dataset')

CLASSIFICATION_DATA_ROOT_DIR = os.path.join(ROOT_DIR, 'Classification')

CLASSIFICATION_TRAIN_DATA_ROOT_DIR = os.path.join(CLASSIFICATION_DATA_ROOT_DIR, 'train')
```

세션 내의 환경에서 실행할 수 있도록 폴더 경로를 각각 /home/work, breast_ultrasound_dataset 로 수정합니다.

Google Drive 마운트해서 데이터셋 가져옴

```
In [ ]: from google.colab import drive
        drive.mount('/content/gdrive/')
Mounted at /content/gdrive/
```

ROOT_DIR, DATA_ROOT_DIR 설정

```
In [ ]: import os

        ROOT_DIR = '/home/work'

        DATA_ROOT_DIR = os.path.join(ROOT_DIR, 'breast_ultrasound_dataset')

        CLASSIFICATION_DATA_ROOT_DIR = os.path.join(ROOT_DIR, 'Classification')

        CLASSIFICATION_TRAIN_DATA_ROOT_DIR = os.path.join(CLASSIFICATION_DATA_ROOT_DIR, 'train')
```

```
In [ ]: import shutil

        if os.path.exists(DATA_ROOT_DIR):
            shutil.rmtree(DATA_ROOT_DIR)

        if os.path.exists(CLASSIFICATION_DATA_ROOT_DIR):
            shutil.rmtree(CLASSIFICATION_DATA_ROOT_DIR)
```

Google Drive 에서 dataset download

```
In [ ]: import shutil
        import os

        try:
            dataset_path = '/content/gdrive/My Drive/Colab Notebooks/dataset'

            shutil.copy(os.path.join(dataset_path, 'Dataset_BUSI_with_GT.zip'), '/content')

        except Exception as err:
            print(str(err))

In [ ]: import zipfile

        with zipfile.ZipFile(os.path.join(ROOT_DIR, 'Dataset_BUSI_with_GT.zip'), 'r') as target_file:
            target_file.extractall(DATA_ROOT_DIR)
```

```
In [ ]: import shutil
        import os
        import glob
```

사진에 표시한 셀들은 실행에 필요하지 않으므로 삭제하겠습니다.



```
In [ ]: import tensorflow as tf
        from tensorflow.keras.layers import Dropout, Dense, GlobalAveragePooling2D
        from tensorflow.keras.applications import MobileNet
```

화면 상단의 재시작 후 모두 실행 아이콘을 클릭하여 모든 셀을 실행합니다.

dense_1 (Dense)	(None, 128)	32896
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 3)	387

```

=====
Total params: 3,524,547
Trainable params: 3,502,659
Non-trainable params: 21,888

```

```
In [22]: model.compile(loss='sparse_categorical_crossentropy', optimizer=tf.keras.optimizers.Adam(2e-5), metrics=['acc'])
```

```
In [*]: hist = model.fit(x_train, y_train, batch_size=32, epochs=50, validation_data=(x_test, y_test))
```

Epoch 1/50

```

2023-03-03 01:20:03.429664: I tensorflow/stream_executor/cuda/cuda_dnn.cc:379] Loaded cuDNN version 8400
2023-03-03 01:20:08.853033: I tensorflow/stream_executor/cuda/cuda_blas.cc:1804] TensorFlow-32 will be used for the matrix mul
tiplication. This will only be logged once.

```

98/98 [=====] - 33s 191ms/step - loss: 1.3364 - acc: 0.4481 - val_loss: 0.9181 - val_acc: 0.5808

Epoch 2/50

98/98 [=====] - 6s 61ms/step - loss: 0.9899 - acc: 0.5744 - val_loss: 0.7869 - val_acc: 0.6744

Epoch 3/50

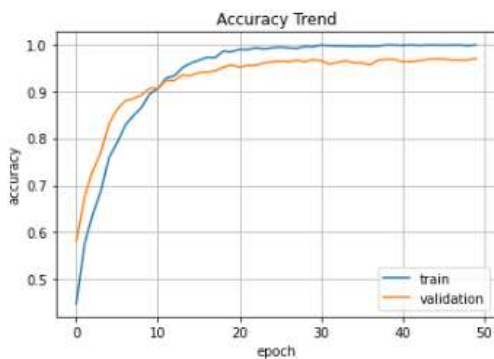
98/98 [=====] - 5s 49ms/step - loss: 0.8188 - acc: 0.6378 - val_loss: 0.6885 - val_acc: 0.7308

Epoch 4/50

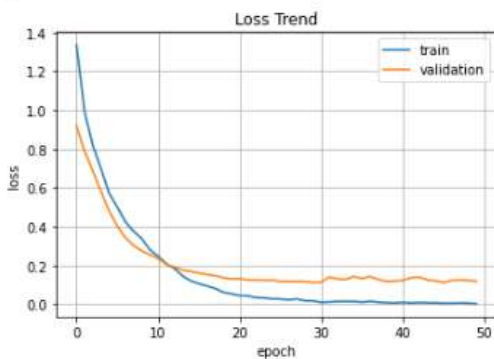
91/98 [=====>...] - ETA: 0s - loss: 0.7017 - acc: 0.6834

모든 셀이 순차적으로 실행되며 학습이 진행되는 것을 확인할 수 있습니다.

```
plt.grid()
plt.show()
```



```
In [25]: plt.plot(hist.history['loss'], label='train')
plt.plot(hist.history['val_loss'], label='validation')
plt.title('Loss Trend')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(loc='best')
plt.grid()
plt.show()
```



```
In [26]: model.evaluate(x_test, y_test)
```

25/25 [=====] - 1s 24ms/step - loss: 0.1195 - acc: 0.9705

Out[26]: [0.11948967725038528, 0.9705128073692322]

학습 결과를 확인합니다.