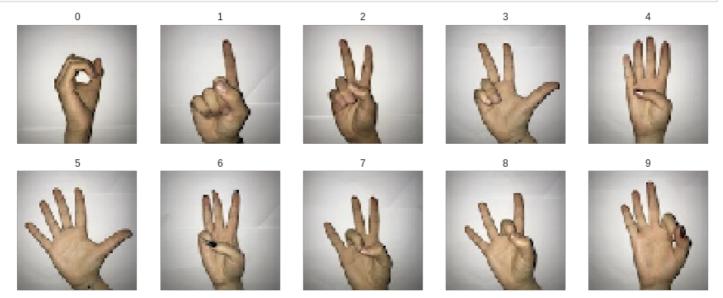
```
In [ ]:
```

```
import os, cv2, math
from keras.models import Sequential
from keras.layers import Conv2D
from keras.layers import MaxPooling2D
from keras.layers import Flatten
from keras.layers import Dense
from keras.layers import Dropout
from keras.preprocessing.image import ImageDataGenerator
from keras.utils import plot model
import numpy as np
from keras.preprocessing import image
from sklearn.model selection import train test split
from shutil import copyfile
from tqdm import tqdm
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
%matplotlib inline
```

In []:

```
nrow, ncol = 2, 5
plt.rcParams['figure.figsize'] = (ncol*3, nrow*3)
for row in range(nrow):
    for col in range(ncol):
        img_index = row*ncol+col
        # load image
        img = image.load_img('Sign-Language-Digits-Dataset/Examples/example_' + str(img_index) + '.JPG', target_size = (64, 64))
        plt.subplot(nrow, ncol, img_index + 1)
        plt.imshow(img)
        plt.title(img_index)
        plt.axis('off')
```



This project aims to create a classifier that can intepret sign language for number 0 to 9. The image of the sign language for respective number is shown above.

Inside 'Sign-Language-Digits-Dataset/Dataset/', the images of different sign language are organised according to their labels(i.e. 0 to 9). The dataset will first be divided into 3 sets: training_set, validation_set, and test_set. The size of training set, validation set, and test set are 70%, 15% and 15% of the whole dataset respectively.

To accomodate for the requirement for flow_from_directory method from keras, I will reorganise the images in the following structure:

- Training set: 'Sign-Language-Digits-Dataset/Dataset/training_set/class_00/image_file'
- Validation set: 'Sign-Language-Digits-Dataset/Dataset/validation set/class 00/image file'

 Test set: 'Sign-Language-Digits-Dataset/Dataset/test_set/class_00/image_file' In []: DATASET PATH = 'Sign-Language-Digits-Dataset/Dataset/' In []: # Creating a list of filename for training set, validation set, and test set train set = {} validation set = {} test set = {} for cat in os.listdir(DATASET PATH): cat dir = os.path.join(DATASET PATH, cat) # e.g. DATASET PATH/'0' cat files = os.listdir(cat dir) # Training set's size is 70% of the data train list , test list = train test split(cat files, test size = 0.3) # Validation set's and Test set's size are both 15% of the data validation_list, test_list = train_test_split(test_list, test_size = 0.5) train set[cat] = train list validation set[cat] = validation list test set[cat] = test list In []: for cat in tqdm(train set.keys()): cat dir = os.path.join(DATASET PATH, 'training set', 'class 0' + str(cat)) os.makedirs(cat dir) for file in train set[cat]: # src path is DATASET PATH/'0'/file src = os.path.join(DATASET PATH, cat, file) # dest path is DATASET PATH/'training set'/'class 00' # to accomodate for the directory format required by flow from directory method in ke ras dest = os.path.join(cat dir, file) copyfile(src, dest) | 10/10 [00:00<00:00, 45.75it/s] 100%| In []: for cat in tqdm(validation set.keys()): cat dir = os.path.join(DATASET PATH, 'validation set', 'class 0' + str(cat)) os.makedirs(cat dir) for file in validation set[cat]: # src path is DATASET PATH/'0'/file src = os.path.join(DATASET PATH, cat, file) # dest path is DATASET PATH/'validation set'/'class 00' # to accomodate for the directory format required by flow from directory method in ke dest = os.path.join(cat dir, file) copyfile(src, dest) | 10/10 [00:00<00:00, 210.44it/s] 100%| In []: for cat in tqdm(test set.keys()): cat dir = os.path.join(DATASET PATH, 'test set', 'class 0' + str(cat)) os.makedirs(cat dir) for file in test set[cat]: # src path is DATASET PATH/'0'/file src = os.path.join(DATASET PATH, cat, file) # dest path is DATASET PATH/'test set'/'class 00'

to accomodate for the directory format required by flow from directory method in ke

ras

dest = os.path.join(cat dir, file)

| 10/10 [00:00<00:00, 208.02it/s]

copyfile(src, dest)

```
In [ ]:
for i in range (10):
 train size = len(train set[str(i)])
  validation size = len(validation set[str(i)])
 test size = len(test set[str(i)])
 print("0{} : Training size({}) Validation size({}) Test size({})".format(i, train size
, validation size, test size))
00 : Training size(143) Validation size(31) Test size(31)
01 : Training size(144) Validation size(31) Test size(31)
02 : Training size(144) Validation size(31) Test size(31)
03: Training size(144) Validation size(31) Test size(31)
04: Training size(144) Validation size(31) Test size(32)
05: Training size(144) Validation size(31) Test size(32)
06: Training size(144) Validation size(31) Test size(32)
07 : Training size(144) Validation size(31) Test size(31)
08 : Training size(145) Validation size(31) Test size(32)
09 : Training size(142) Validation size(31) Test size(31)
Data augmentation is performed on the training set images so that the classifier can learn any changes with
respect to scaling, horizontal_flip, or others.
In [ ]:
```

```
# Performing data augmentation on training dataset
train datagen = ImageDataGenerator(rescale = 1./255,
                                   shear range = 0.2,
                                   zoom range = 0.2,
                                   horizontal flip = True)
# For validation dataset, only rescale the pictures
validation datagen = ImageDataGenerator(rescale = 1./255)
# For test dataset, only rescale the pictures
test datagen = ImageDataGenerator(rescale = 1./255)
training data = train datagen.flow from directory(os.path.join(DATASET PATH, 'training se
t'),
                                                 target size = (64, 64),
                                                 batch size = 32,
                                                 class mode = 'categorical')
validation data = validation datagen.flow from directory(os.path.join(DATASET PATH, 'vali
dation set'),
                                                 target size = (64, 64),
                                                 batch size = 32,
                                                 class mode = 'categorical')
test_data = test_datagen.flow_from_directory(os.path.join(DATASET_PATH, 'test_set'),
                                            target size = (64, 64),
                                            batch size = 32,
                                            class mode = 'categorical')
```

Found 310 images belonging to 10 classes. Found 310 images belonging to 10 classes. Found 314 images belonging to 10 classes.

In []:

```
# Initialising the CNN
classifier = Sequential()

# Adding first convolutional layer, followed by pooling, and dropout
classifier.add(Conv2D(32, (3, 3), input_shape = (64, 64, 3), activation = 'relu'))
classifier.add(MaxPooling2D(pool_size = (2, 2)))
classifier.add(Dropout(0.25))

# Adding second convolutional layer, followed by pooling, and dropout
classifier.add(Conv2D(32, (3, 3), activation = 'relu'))
```

```
classifier.add(MaxPooling2D(pool size = (2, 2)))
classifier.add(Dropout(0.25))
# Adding third convolutional layer, followed by pooling, and dropout
classifier.add(Conv2D(32, (3, 3), activation = 'relu'))
classifier.add(MaxPooling2D(pool size = (2, 2)))
classifier.add(Dropout(0.25))
# Flattening
classifier.add(Flatten())
# Full connection
classifier.add(Dense(units = 128, activation = 'relu'))
classifier.add(Dense(units = 10, activation = 'softmax'))
# Compiling the CNN
classifier.compile(optimizer = 'adam', loss = 'categorical_crossentropy', metrics = ['ac
curacy'])
In [ ]:
# Train the data with training set, and check the result with validation accuracy
history = classifier.fit_generator(training_data,
               steps per epoch = math.ceil(training data.n / training data.bat
ch size),
               epochs = 100,
               validation data = validation data,
               validation steps = math.ceil(validation data.n / validation dat
a.batch size))
Epoch 1/100
loss: 2.2999 - val acc: 0.2000
Epoch 2/100
loss: 2.2814 - val acc: 0.2032
Epoch 3/100
loss: 1.7885 - val acc: 0.4677
Epoch 4/100
loss: 1.3537 - val acc: 0.5968
Epoch 5/100
loss: 1.1175 - val acc: 0.6548
Epoch 6/100
- val acc: 0.7548
Epoch 7/100
loss: 0.7076 - val acc: 0.7645
Epoch 8/100
```

1/45 [.....] - ETA: 1s - loss: 0.6897 - acc: 0.750045/45 [==== ==========] - 4s 85ms/step - loss: 0.5160 - acc: 0.8338 - val loss: 0.3950

loss: 0.6002 - val acc: 0.8161

loss: 0.5330 - val acc: 0.8323

loss: 0.4652 - val acc: 0.8387

loss: 0.4387 - val acc: 0.8581

loss: 0.3751 - val acc: 0.8710

Epoch 9/100

Epoch 10/100

Epoch 11/100

Epoch 12/100

Epoch 14/100

- val_acc: 0.8742
Epoch 13/100

```
loss: 0.3411 - val acc: 0.8871
Epoch 15/100
loss: 0.3274 - val acc: 0.8968
Epoch 16/100
loss: 0.2720 - val acc: 0.9097
Epoch 17/100
- val acc: 0.9194
Epoch 18/100
loss: 0.2637 - val acc: 0.9129
Epoch 19/100
loss: 0.2287 - val acc: 0.9194
Epoch 20/100
loss: 0.2743 - val acc: 0.9065
Epoch 21/100
loss: 0.3163 - val acc: 0.9032
Epoch 22/100
loss: 0.2428 - val acc: 0.9194
Epoch 23/100
loss: 0.2146 - val acc: 0.9355
Epoch 24/100
loss: 0.2397 - val acc: 0.9226
Epoch 25/100
loss: 0.1783 - val acc: 0.9387
Epoch 26/100
loss: 0.1967 - val acc: 0.9290
Epoch 27/100
loss: 0.2683 - val_acc: 0.9065
Epoch 28/100
35/45 [============>.....] - ETA: 0s - loss: 0.1885 - acc: 0.931345/45 [====
- val acc: 0.9452
Epoch 29/100
loss: 0.2098 - val acc: 0.9290
Epoch 30/100
loss: 0.1798 - val acc: 0.9484
Epoch 31/100
loss: 0.2427 - val acc: 0.9226
Epoch 32/100
loss: 0.2270 - val acc: 0.9387
Epoch 33/100
loss: 0.2692 - val acc: 0.9226
Epoch 34/100
loss: 0.2032 - val acc: 0.9419
Epoch 35/100
loss: 0.1822 - val acc: 0.9516
Epoch 36/100
loss: 0.1751 - val acc: 0.9452
Epoch 37/100
```

```
loss: 0.1665 - val_acc: 0.9484
Epoch 38/100
loss: 0.1777 - val acc: 0.9548
Epoch 39/100
========= ] - 4s 86ms/step - loss: 0.1319 - acc: 0.9590 - val_loss: 0.1698
- val acc: 0.9710
Epoch 40/100
loss: 0.1491 - val acc: 0.9581
Epoch 41/100
loss: 0.1482 - val acc: 0.9645
Epoch 42/100
loss: 0.2051 - val acc: 0.9290
Epoch 43/100
loss: 0.1704 - val acc: 0.9516
Epoch 44/100
loss: 0.2136 - val acc: 0.9323
Epoch 45/100
4/45 [=>.....] - ETA: 0s - loss: 0.1298 - acc: 0.945345/45 [=====
- val acc: 0.9387
Epoch 46/100
loss: 0.1994 - val acc: 0.9355
Epoch 47/100
loss: 0.2190 - val acc: 0.9484
Epoch 48/100
loss: 0.1633 - val acc: 0.9548
Epoch 49/100
loss: 0.1364 - val acc: 0.9613
Epoch 50/100
- val acc: 0.9419
Epoch 51/100
loss: 0.1538 - val acc: 0.9774
Epoch 52/100
loss: 0.2437 - val acc: 0.9323
Epoch 53/100
loss: 0.2176 - val acc: 0.9484
Epoch 54/100
loss: 0.2447 - val acc: 0.9258
Epoch 55/100
loss: 0.1845 - val acc: 0.9677
Epoch 56/100
1/45 [......] - ETA: 1s - loss: 0.0228 - acc: 1.000045/45 [====
- val acc: 0.9677
Epoch 57/100
loss: 0.1540 - val acc: 0.9645
Epoch 58/100
loss: 0.2367 - val acc: 0.9452
Epoch 59/100
loss: 0.2098 - val acc: 0.9516
Epoch 60/100
```

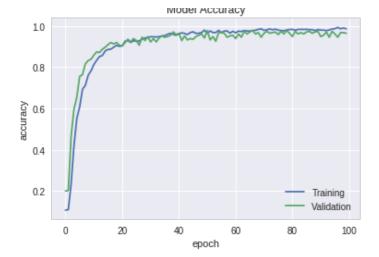
```
loss: 0.1907 - val acc: 0.9548
Epoch 61/100
========== ] - 4s 87ms/step - loss: 0.0942 - acc: 0.9666 - val loss: 0.2337
- val acc: 0.9387
Epoch 62/100
loss: 0.1878 - val acc: 0.9613
Epoch 63/100
loss: 0.2179 - val acc: 0.9452
Epoch 64/100
loss: 0.1301 - val acc: 0.9742
Epoch 65/100
loss: 0.1568 - val acc: 0.9613
Epoch 66/100
loss: 0.1773 - val acc: 0.9710
Epoch 67/100
6/45 [===>.....] - ETA: 1s - loss: 0.0278 - acc: 0.989645/45 [====
======== ] - 4s 87ms/step - loss: 0.0618 - acc: 0.9778 - val_loss: 0.1486
- val acc: 0.9774
Epoch 68/100
loss: 0.1406 - val acc: 0.9613
Epoch 69/100
loss: 0.1386 - val acc: 0.9677
Epoch 70/100
loss: 0.2163 - val acc: 0.9452
Epoch 71/100
loss: 0.1571 - val acc: 0.9645
Epoch 72/100
- val_acc: 0.9742
Epoch 73/100
loss: 0.1564 - val acc: 0.9645
Epoch 74/100
loss: 0.1113 - val acc: 0.9677
Epoch 75/100
loss: 0.1691 - val acc: 0.9710
Epoch 76/100
loss: 0.1485 - val acc: 0.9581
Epoch 77/100
loss: 0.1324 - val acc: 0.9742
Epoch 78/100
6/45 [===>.....] - ETA: 1s - loss: 0.0139 - acc: 0.994845/45 [====
- val acc: 0.9613
Epoch 79/100
loss: 0.1371 - val acc: 0.9774
Epoch 80/100
loss: 0.1531 - val acc: 0.9645
Epoch 81/100
loss: 0.1479 - val acc: 0.9484
Epoch 82/100
loss: 0.1519 - val acc: 0.9742
```

```
Epoch 83/100
- val acc: 0.9613
Epoch 84/100
loss: 0.1799 - val acc: 0.9677
Epoch 85/100
loss: 0.1530 - val acc: 0.9613
Epoch 86/100
loss: 0.1274 - val acc: 0.9710
Epoch 87/100
loss: 0.1294 - val acc: 0.9742
Epoch 88/100
loss: 0.1713 - val acc: 0.9645
Epoch 89/100
4/45 [=>.....] - ETA: 0s - loss: 0.0368 - acc: 0.976645/45 [=====
- val acc: 0.9710
Epoch 90/100
loss: 0.1285 - val acc: 0.9742
Epoch 91/100
loss: 0.1770 - val acc: 0.9484
Epoch 92/100
loss: 0.1877 - val acc: 0.9548
Epoch 93/100
loss: 0.1339 - val acc: 0.9710
Epoch 94/100
- val_acc: 0.9452
Epoch 95/100
loss: 0.1371 - val acc: 0.9742
Epoch 96/100
loss: 0.1387 - val acc: 0.9613
Epoch 97/100
loss: 0.2287 - val acc: 0.9452
Epoch 98/100
loss: 0.1480 - val acc: 0.9677
Epoch 99/100
loss: 0.1255 - val acc: 0.9677
Epoch 100/100
1/45 [.....] - ETA: 1s - loss: 0.0529 - acc: 0.968845/45 [====
- val acc: 0.9645
In [ ]:
plt.plot(history.history['acc'])
plt.plot(history.history['val acc'])
plt.title('Model Accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['Training', 'Validation'])
```

<matplotlib.legend.Legend at 0x7f0bb8f88748>

Out[]:

Madal Assurance



From the accuracy plot, the validation accuracy differs from the training accuracy by a small extent indicating absence of overfitting. When the classifier is evaluated on the test set, it obtains a relatively high accuracy.

In []:

```
# Accuracy of the classifier when evaluated based on the test_set
test_loss, test_accuracy = classifier.evaluate_generator(test_data, math.ceil(test_data.n
/ test_data.batch_size))
print("Accuracy on test set : {}".format(test_accuracy))
```

Accuracy on test set : 0.974522290715746

To visualise the performance of the classifier, the classifier will be used to predict all the example images for 0 to 9.

In []:

```
nrow, ncol = 2, 5
plt.rcParams['figure.figsize'] = (ncol*3, nrow*3)
for row in range(nrow):
   for col in range(ncol):
       img_index = row*ncol+col
        # load image
       img = image.load img('Sign-Language-Digits-Dataset/Examples/example ' + str(img
index) + '.JPG', target_size = (64, 64))
        # convert image into array for prediction
       test_image = image.img_to_array(img)
        test_image = np.expand_dims(test_image, axis = 0)
        # predict image using classifier
       result = classifier.predict(test image).argmax()
       plt.subplot(nrow, ncol, img index + 1)
       plt.imshow(img)
       plt.title("Actual({}) Predicted({})".format(img index, result))
       plt.axis('off')
```



Actual(5) Predicted(5)

















It can be seen from above output that the classifier is able to classify all the images correctly. To save the model for future use, simply run the code below.

```
In [ ]:
```

```
# save the models and weight for future purposes
# serialize model to JSON
model_json = classifier.to_json()
with open("model.json", "w") as json_file:
        json_file.write(model_json)
# serialize weights to HDF5
classifier.save_weights("model.h5")
print("Saved model to disk")
```

Saved model to disk

Reference

- https://github.com/ardamavi/Sign-Language-Digits-Dataset
- https://www.superdatascience.com/deep-learning/