19BCE248

Big Data Analytics

Lab 2

D2

AIM: To apply Machine Learning to any bigger dataset and to analyze the limitation of it.

Dataset: Cifar-10 dataset used for classification of images in 10 classes namely airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.

The dataset is divided into five training batches and one test batch, each with 10000 images. The test batch contains exactly 1000 randomly-selected images from each class. The training batches contain the remaining images in random order, but some training batches may contain more images from one class than another. Between them, the training batches contain exactly 5000 images from each class.

Dataset Size: Approx 6-7GB of image data

Model Used: Transfer Learning

```
model = Sequential()
model.add(xcep)
model.add(GlobalAveragePooling2D())
model.add(Dropout(0.3))
model.add(Dense(10, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

With inbuilt Xception pre-defined model in keras and adding above layers

Limitations:

Time: As seen that just to run 5 epochs we need approx. 600 sec which leads to 0.5 epoch/sec which is not acceptable in real life as we need to have faster response. These can be due to huge amount of data which leads to use of Big Data

```
Epoch 1/5
2022-08-24 09:56:48.606547: I tensorflow/stream_executor/cuda/cuda_dnn.cc:369] Loaded cuDNN version 8005
332/332 [==============================] - 136s 342ms/step - loss: 0.5386 - accuracy: 0.8221 - val_loss: 0.5012
- val_accuracy: 0.8357
Epoch 2/5
332/332 [==============================] - 116s 341ms/step - loss: 0.2722 - accuracy: 0.9096 - val_loss: 0.5463
- val_accuracy: 0.8358
Epoch 3/5
332/332 [==============================] - 119s 351ms/step - loss: 0.2012 - accuracy: 0.9333 - val_loss: 0.3208
- val_accuracy: 0.8973
Epoch 4/5
332/332 [==============================] - 118s 345ms/step - loss: 0.1614 - accuracy: 0.9455 - val_loss: 0.3011
- val_accuracy: 0.9012
Epoch 5/5
332/332 [==============================] - 115s 337ms/step - loss: 0.1363 - accuracy: 0.9539 - val_loss: 0.3396
- val_accuracy: 0.8951
seconds passed: 605.4247193336487
```

Kaggle link

https://www.kaggle.com/code/dhruvil01/prac3-transfer-learning