Verzeo Major Project- DS-MAJOR-FEB

Project Title- Indian Human Image Classification Model

Submitted by- Agrata Dwivedi

Group number – B14

Respected Sir/Ma'am ,this is Agrata Dwivedi from group B14 , I am going to start with our project strategy. So, firstly starting with the data collection part here,initially we searched for images of Indian human with a particular skin tone from google,bing, websites such as getty images. We used an online image scrapper named Parsehub from where we were able to download the images of the URL in a CSV file and extracted them in a specific folder. We had stored all the images in a specific folder . For Non-Indian data, a dataset from Kaggle helped us to collect almost 600 images .After this part we divided the category of Indian Human into certain subcategories as Politicians , Sportsperson , children ,north -eastern states,actors,actresses ,etc.For non-human category almost 500 images were collected which included images of objects,cars,animals,plants etc .

Now comes the Data Refining part of our project.Firstly the division was between Human and Non Human then the human folder is further divided into 2 categories – Indian and Non Indian .Indian images folder would contain images for all skin types ,age and sex . After collection half part of the data is manually refined and other half part is refined using open CV image classification and CNN -Convolution Neural Network .Now we come to our third step of data processing here, all images are read and converted into array by appending using for loop . The same process is followed for all the labels .The set is then divided into testing and training data .Image augmentation using CNN was applied to increase the range of the model .

Now, the last step is model building which includes specifying architecture,compiling,fitting and predicting This is a classification model since we have fixed number of inputs . CNN was used instead of normal neural network for increasing scope and functionality of code.

Project Description-

- It was an image classification model where we have given an image input and our model would classify it into Indian human from rest of the world.
- The dataset of our group comprised of 100 images of humans with different skin tones , human images that originate from India vs others also we collected images of non-humans which included plants,animals,objects etc.Further we used Open CV library

Open CV- It was used to download the images and convert them into numpy array. Since it was a multi-dimensional picture it consists of multidimensional array which had 4 indexes .First how many pictures and second and third would be dimensions . For eg- 100x100 pixels and fourth one consists of color channels like if it's 1 then it's grey scale and if 3 then there are 3 values . So, mostly colored pictures are of 3 colour channels which are blue,red and green.

- Now we split the images into testing and training data also we added labels such as 0 and 1
- This labeling of images states that if the image is non human then the label would be 0 and vice- versa .
- Thus,in our project labels are our targets and all the images are our features.
- We had made some changes like we used image- data generator to create different kinds of images from the training images and then we divided into our two batches and finally we applied our Sequential model .
- Sequential model is basically an subset of CNN. There are various kinds of models in CNN and sequential model is one of it's kind.

- Then we added layers of convolution 2D and dropout flatten and after that we fitted our model after compiling it.
- Now we specified the epochs like if the epoch is 6 then the model is fitted 6 times.
- After this we stored it in a folder named results either in desktop or any location .
- Then we specified the path that our model with predict all the images and then if the prediction is greater than a particular value it means that it could be non human or non Indian
- The final accuracy we achieved was of 78.6 %

My Individual Contribution-

My contribution in this project was that I have collected images using websites such as getty images and used an online image scrapper named Parsehub to download the images. I had successfully submitted all the recorded videos of our tasks in the google form . I used OBS studio for recording the work . I have helped my team members in evaluating the code and was successful in increasing the accuracy.

So at last I would like to thank the entire team of Verzeo for giving me this opportunity to interact and coordinate in a team and it was only possible because of Verzeo that I learned new skills from this entire journey.

References -

How to Scrape and Download ALL images from a webpage with Python

https://www.kaggle.com/ashwingupta3012/human-faces

https://www.analyticsvidhya.com/blog/2019/01/build-image-classification-model-10-minutes/?utm_source=blog&utm_medium=multi-label-image-classification

https://www.google.com/url?sa=t&source=web&rct=j&url=https://medium.com/analytics-vidhya/export-images-from-jupyter-notebook-with-a-single-command-422db2b66e92&ved=2ahUKEwi5hK2DvP_vAhWVbysKHSGqA7AQFjABegQIAxAF&usg=AOvVaw3SG7Gko13Owg1Wpljrwp-Y&cshid=1618462916621

Tutorial 24 - Saving images from python to your local drive