**Dataset Information**

We have collected images from 5 different datasets all of which are publicly available.

* Dataset 1: <https://www.kaggle.com/omkargurav/face-mask-dataset>

Face Mask Detection Dataset by *Omkar Gurav*

* Dataset 2: <https://www.kaggle.com/prithwirajmitra/covid-face-mask-detection-dataset>

COVID Face Mask Detection Dataset by *Prithwiraj Mitra*

* Dataset 3: <https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset>

Face Mask ~12k Images Dataset by *Ashish Jangra*

* Dataset 4: <http://www.vision.caltech.edu/pmoreels/Datasets/Home_Objects_06/> Home Objects dataset by *Caltech*
* Dataset 5: <https://www.kaggle.com/prasunroy/natural-images>

Natural Images by *Prasun Roy*

The composition of training data is as follows:

* **Mask:** 1100 images
  + 400 images from folder ‘with\_mask’ in Dataset 1 (1-100, 1001-1100, 2501-2700)
  + All 300 images from ‘Mask’ category in ‘Train’ folder of Dataset 2
  + 400 images from ‘WithMask’ category in ‘Train’ folder of Dataset 3 (4-1011)
* **Without Mask:** 1100 images
  + 400 from folder ‘without\_mask’ in Dataset 1 (1-100, 1001-1100, 2501-2700)
  + All 300 images from ‘Non Mask’ category in ‘Train’ folder of Dataset 2
  + 400 images from ‘WithoutMask’ category in ‘Train’ folder of Dataset 3 (1501-1732,3001-3234)
* **Not a Person:** 1150 images
  + 50 images from directory ‘Training images’ of Dataset 4 (P1020583-P1020772)
  + 1100 images from different folders of Dataset 5 (0-99 from airplane, car, flower, fruit, motorbike and 0-299 from cat, dog)

The composition of testing data is as follows:

* **Mask:** 200 images
  + 75 images from folder ‘with\_mask’ in Dataset 1 (2531-2605)
  + All 50 images from ‘Mask’ category in ‘Test’ folder of Dataset 2
  + 75 images from ‘WithMask’ category in ‘Test’ folder of Dataset 3 (3-1820)
* **Without Mask:** 200 images
  + 75 from folder ‘without\_mask’ in Dataset 1 (601-675)
  + All 50 images from ‘Non Mask’ category in ‘Test’ folder of Dataset 2
  + 75 images from ‘WithoutMask’ category in ‘Test’ folder of Dataset 3 (1-754)
* **Not a Person:** 200 images
  + 45 images from directory ‘Training images’ of Dataset 4 (P1020773-P1030158)
  + 155 images from different folders of Dataset 5 (201-220 from airplane, car, flower, 201-216 from fruit and 401-440 from cat, dog)

Size of the dataset- 200MB