AI Mask Detector

COMP 6721 Applied Artificial Intelligence (Fall 2021)
Project Assignment

Team: AK_02

Akanksha Panwar
Data Specialist
40139125

Jatin Katyal
Training Specialist
40196189

Rucha Shende
Evaluation Specialist
40205356

Contents

List of Tables

List of Figures

1 Dataset

It is very crucial to carefully select the data that is relevant to the application. Considering the qualitative aspect, it is best to rely on datasets that are currently available and widely used by peers. For the quantity aspect, even though there are techniques like Classification from Single Example[3] and One-Shot Learning[2] that can perform equally well on tasks when compared to conventional methods, we preferred to create a sufficiently large pool of data for the sake of simplicity.

1.1 Publicly Available Data

We decided to re-purpose the datasets that were meant to be used in a specific application but contained the information that we are looking for, since no dataset can be readily used. Our aim was to collect maximum possible images of one or more subjects wearing a facial mask (Cloth, FFP2 or Surgical).

From Kaggle, we selected the Face Mask Dataset[5], the data set has 853 images of people wearing different kinds of masks with one or more person in the frame. Using the bounding boxes for each face provided in the annotations, 4000 faces from the original 853 images were extracted. The extracted images were then manually labeled as Cloth Mask, FFP2 Mask, No Mask & Surgical Mask. Many images were dropped because of small in resolution or occlusion, resulting in 728 images which were unevenly spread across the classes. Later, we found two more datasets on Kaggle that have more images that suit our requirements [6] [4]. These new datasets have combined thirteen thousand images of masked and unmasked people but, since we already had a big chunk of data gathered already we only acquired 334 images from this collection.

1.2 Data Sourced from Internet

After carefully selecting the images from above dataset we found a class imbalance in our dataset and we could use more images for the task. To supplement our dataset we searched for images related our task on the internet which lead to many websites[1]. Combined 264 images were collected from various websites such that all classes have equal distribution of data in final dataset. Table 1 shows the near equal distribution of classes.

Class	Count
Cloth Mask	346
FFP2 Mask	329
No Mask	333
Surgical Mask	325

Table 1: Class distribution of images.

2 Architecture

2.1 Model

The model trained for detecting different masks uses six convolution layers and 3 max-pool layers in such a fashion that after every two convolution layers a max-pool layer is used. Each convolution layer uses a filter of size 3 and stride 1, we also used padding of size 1 to maintain the resolution after every convolution. Each max-pool layer uses a filter of size 2 and stride of size 2. Convolution and max-pooling layers are used in such a way that after every 2 convolutions, resolution the intermediate feature representation is reduced by half and the number of channels are doubled. The idea of using 2 consecutive layers of smaller size to have an effective field of larger filter is quite old but, inspired from widely successful architecture introduced by K. Simonyan A. Zisserman in 2014[8]. After all convolutions the features are flattened and passed through a fully connected network which has 50 hidden units between the results of convolution and final layer of 4 neurons. Figure 1 shows the same architecture.

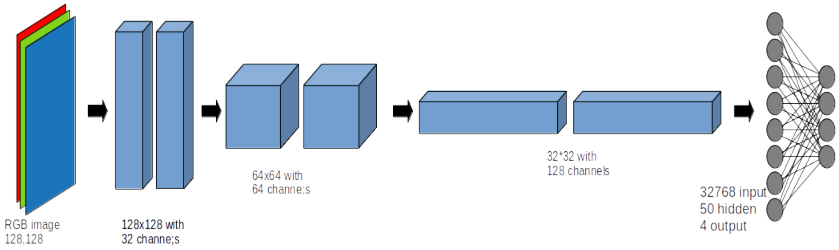

Figure 1: Architecture of the model trained for the task.

2.2 Training

For training we used Adam optimizer and Cross Entropy loss, Adam optimizing algorithm is computationally efficient and uses both momentum and RMSProp and thus is better than vanilla Stochastic Gradient Descent[7]. Model was trained on mini-batches of size 32, with a learning rate of 0.00001 for 15 epochs. Training was done on Nvidia Geforce GTX 1660 TI GPU with CUDA.

3 Evaluation

3.1 Data Split

The entire data was split into 3 subsets, namely training set, validation set and testing set. Twenty percent data was reserved for testing at the beginning, remaining eighty percent data was further split. Twenty percent of the remaining data was retained for validation and rest was used for training. We pragmatically created the splits within the code and assigned a data loader instance to each split and used methods from sklearn package to get results shown in Table 2. Additionally, we also provided the multi-class confusion matrix in Table 3.

Metric	Train Set	Validation Set	Test Set
Accuracy	92.93	71.70	73.21
Precision	93.63	76.01	74.79
Recall	93.93	71.70	73.21
f1	92.92	70.75	73.25

Table 2: Values for different metrics after training

	Actual			
Predicted	Cloth Mask	FFP2	No Mask	Surgical Mask
Cloth Mask	37	5	12	18
FFP2 Mask	3	25	1	21
No Mask	4	0	65	1
Surgical Mask	11	5	0	57

Table 3: Confusion matrix for test set after training

3.2 Improvements

We achieved a test accuracy 73% without much optimization, we expect to further increase this measure by introducing weight initialization and better hyper parameter optimization. We already explored a few frontiers by changing the number of consecutive layers, total number of convolution layers, number of hidden units in fully connected layer and number of hidden layers in fully layers.

References

[1] https://www.google.com https://www.dawn.com
https://www.healthgrades.com https://www.stocksy.com
https://www.india.com https://www.henryford.com
https://www.bigstockphoto.com https://www.kcallergyasthma.com
https://www.alamy.com https://carithersgroup.com
https://www.istockphoto.com https://zeenews.india.com
https://www.cnn.com https://www.gq.com
https://www.rediff.com https://elements.envato.com
https://www.livescience.com https://gadgets.ndtv.com
https://www.ucsf.edu https://www.gettyimages.co.uk
https://economictimes.indiatimes.com https://health.clevelandclinic.org
https://www.freepik.com https://www.youtube.com
https://photodune.net https://weather.com
https://www.eehealth.org https://news.abplive.com
https://borgenproject.org https://www.nytimes.com
https://abcnews.go.com https://www.bbc.com
https://unsplash.com https://www.hindustantimes.com
https://www.shutterstock.com https://www.sfgate.com
https://www.latimes.com https://www.picxy.com .

[2] Li Fei-Fei, R. Fergus, and P. Perona. “One-shot learning of object categories”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.4 (2006), pp. 594–611. DOI: 10.1109/TPAMI.2006.79.

[3] Michael Fink. “Object Classification from a Single Example Utilizing Class Relevance Metrics.” In: Jan. 2004.

[4] *Kaggle, Covid Face Mask Detection Dataset.* URL: https://www.kaggle.com/prithwirajmitra/covid-face-mask-detection-dataset.

[5] *Kaggle, Face Mask Detection.* URL: https://www.kaggle.com/andrewmvd/face-mask-detection.

[6] *Kaggle, Face Mask Detection 12K Images Dataset.* URL: https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset.

[7] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization.* 2017. arXiv: 1412.6980 [cs.LG].

[8] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition.* 2015. arXiv: 1409.1556 [cs.CV].