Objective

Train a model to detect images with scratches on the text.

Dataset Link

1. Non-zipped folder
   https://drive.google.com/drive/folders/1jijQD_HSRffKjLfw2z7IFekbGzN-l7pq?usp=drive_link

2. Zipped folder
   https://drive.google.com/file/d/187aiGbDYc14T1PW-Lx7IHqvdLflc31a9/view?usp=sharing

Please request viewing access to whichever link you are using.

Note that this is private data. Only certain people have access to it. Therefore, do not share it publicly.

Details

1. In the dataset link, you will find ‘good’ images with clear text and ‘bad’ images with scratched text.
2. During the test, an input image will be given, and the model must predict whether the image has a scratch on the text.
3. The dataset's images contain three types of text. The test dataset will also contain images with the same text.
4. Bonus points will be given if the model can also create a bounding box or mask over the scratch.

Judging criteria

1. How many approaches have you explored and tried to solve the problem? The more approaches, the better, especially if you can’t get any approach to work perfectly.
2. The recall value of the model for bad images, i.e., no bad images, should be left out.
3. The precision of images classified as bad images i.e., minimization of false positives (positive being classification of bad image)
4. Bonus points will be given
    a. if your model can also create a bounding box or mask on the scratch in the bad image

b. If you allow the user to control the threshold of the size of the scratch, to classify the image as a bad image.
c. If you can figure out ways to increase the dataset size of bad images beyond the typical data augmentation techniques.
d. If the model could be extended to scratches on other surfaces, such as metallic surfaces, phone screens etc.

Submission

1. Create a PRIVATE Github repo. Give access to puru07 and add the following to it -
    a. The scripts and all the trained weights.
    b. Readme on how to try out your best models.
2. A Google doc/ pdf containing
    a. details of the approaches you tried
    b. Results of each approach
    c. Why do you think it worked or didn't work?