Cut Mix

13.12.2020

Algorithm

In each epoch, a random number is sampled from a uniform distribution. If the number is smaller than a chosen threshold t, for each image, a randomly sampled smaller patch of it is replaced by the same patch area from another image in the batch.
If the sampled number is greater than the threshold t, the batch goes into the model without augmentation.
For more details, kindly refer to Section 3.1 of the paper attached in the same email.

Experiments

1)
Train set size = 897(False) + 463(True)
Validation set size = 231(True) + 231(False)
Test set size = 174(True) + 174(False)you

Trained for 100 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.5
Optimizer = Adam
FIxed learning rate = 0.001

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	73.27%	74.08%	70.11%	71.89%	76.43%
Fold2	79.02%	86.3%	68.9%	74.16%	89.89%
Fold3	73.5%	73.56%	73.56%	73.56%	73.56%

2)
Train set size = 897(False) + 463(True)
Validation set size = 231(True) + 231(False)
Test set size = 174(True) + 174(False)

Trained for 100 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.6
Optimizer = Adam
Fixed Learning rate = 0.001

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	72.98%	74.69%	69.5%	71.51%	76.4%
Fold2	78.19%	79.89%	75.28%	76.6%	81.03%
Fold3	74.13%	73.33%	75.86%	75%	72.4%

3)
Train set size = 897(False) + 463(True)
Validation set size = 231(True) + 231(False)
Test set size = 174(True) + 174(False)

Trained for 300 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.5
Optimizer = Adam
Fixed Learning rate = 0.001

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	74.71%	79.05%	67.24%	71.5%	82.18%
Fold2	82.18%	85.44%	77.5%	79.47%	86.78%
Fold3	75%	82.7%	63.2%	70.2%	86.78%

4)
Train set size = 897(False) + 463(True)
Validation set size = 231(True) + 231(False)
Test set size = 174(True) + 174(False)

Trained for 300 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.6
Optimizer = Adam
Fixed Learning rate = 0.001

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	73.2%	72.62%	74.71%	73.96%	71.83%
Fold2	74.71%	73.36%	77.5%	76.2%	71.83%
Fold3	74.13%	74.7%	72.8%	73.59%	75.28%

5)
Train set size = 897(False) + 463(True)
Validation set size = 231(True) + 231(False)
Test set size = 174(True) + 174(False)

Trained for 300 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.5
Optimizer = Adam
Initial learning rate = 0.1, decayed by a factor of 0.1 at epochs 75, 150 and 225

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	50%	*	0%	50%	100%
Fold2	50%	*	0%	50%	100%
Fold3	71.5%	79.5%	58.04%	66.96%	100%

* Division by Zero error. Model predicted all samples to be negative because of which denominator resulted in zero in accordance with the below equation.

$$Precision = \frac{\#True\ Positives}{\#True\ Positives + \#False\ Positives}$$

6)
Train set size = 897(False) + 463(True)
Validation set size = 231(True) + 231(False)
Test set size = 174(True) + 174(False)

Trained for 300 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.5

Optimizer = Adam
Initial learning rate = 0.01, decayed by a factor of 0.1 at epochs 75, 150 and 225

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	82.4%	87.41%	75.86%	78.6%	89.08%
Fold2	77.8%	92.17%	60.9%	70.8%	94.8%
Fold3	83.33%	84.11%	82.18%	82.5%	84.48%

7)
Train set size = 1497(False) + 1063(True)
Validation set size = 231(True) + 231(False)
Test set size = 174(True) + 174(False)

Trained for 300 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.5
Optimizer = Adam
Initial learning rate = 0.01, decayed by a factor of 0.1 at epochs 75, 150 and 225

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	82.4%	87.41%	75.86%	78.6%	89.08%
Fold2	80.7%	92.12%	67.24%	74.2%	94.25%
Fold3	83.6%	89.2%	76.4%	79.3%	90.8%

8)
Train set size = 1497(False) + 1063(True)
Validation set size = 231(True) + 231(False)

Test set size = 174(True) + 174(False)

Trained for 400 epochs, BatchSize = 34
Used bicubic interpolation for resizing to 256X256
Threshold t = 0.5
Optimizer = Adam
Initial learning rate = 0.01, decayed by a factor of 0.1 at epochs 75, 150, 225 and 300

	Accuracy	Defective Precision	Defective Recall	Non Defective Precision	Non Defective Recall
Fold1	78.4%	74.8%	85.6%	83.2%	71.26%
Fold2	80.17%	83.43%	75.28%	77.4%	85%
Fold3	79.3%	77.71%	82.18%%	87.05%	76.4%

Cutmix does seem to have imparted some stability. In the experiments conducted prior to incorporating cutmix, there were instances where the metrics' values collapsed to ~20% or 30%. So far, such an instance has not occurred in the experiments using cutmix.