BUSI preprocessing & augmentation

* BUSI Dataset size:
* Malignant: 210 images and 211 masks
* Normal: 133 images and 133 masks
* Benign: 437 images and 454 masks
* Total: 780 images and 798 masks

Step 1

* 1. Data preprocessing for full images:
* Remove TOP LEFT annotation
* Letters, texts, drawings removal
* Contrast enhancement
* Resizing the image to the target size (224,224)
  1. Data preprocessing for mask images:
* Resizing the image to the target size (224,224)
* BUSI Dataset size after 1.1 and 1.2:
* Malignant: 210 images and 211 masks
* Normal: 133 images and 133 masks
* Benign: 437 images and 454 masks
* Total: 780 images and 798 masks

Step 2

1. Dataset creation for each label (Malignant, Normal, Benign):

* Overlapping of multiple masks in one image, for each image, so now :

number of images = number of masks

* BUSI Dataset size after 2:
* Malignant: 210 images and 210 masks
* Normal: 133 images and 133 masks
* Benign: 437 images and 437 masks
* Total: 780 images and 780 masks

Step 3

1. Dataset train-validation-test splitting (per class):

* Train: 70%
* Validation: 15%
* Test: 15%

* BUSI Dataset size after 3:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Malignant | | | Normal | | | Benign | | |
| Train | Validation | Test | Train | Validation | Test | Train | Validation | Test |
| 147 | 31 | 32 | 93 | 20 | 20 | 305 | 66 | 66 |

Step 4

1. Data augmentation:

* Flip
* Rotate
* Zoom
* Noise
* Brightness
* These operation are done on train data
* BUSI Dataset after step 4:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Malignant | | | Normal | | | Benign | | |
| Train | Validation | Test | Train | Validation | Test | Train | Validation | Test |
| 588  Total: | 31 | 32 | 558 | 20 | 20 | 610 | 66 | 66 |
| 1756 | | | | | | | | |