**Objective**: **in general**: lung cancer diagnosis. **specifically**, early detection for lung cancer

**Experiments:**  *framing into two approach with a fixed preprocessing steps*

* **Data preprocessing** Normalization, image resizing ,Then split data it into train and test
* **Approach one:** Preprocessed data go to the custom CNN architecture to train.

A chart of different colored squares

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| Batch size | Loss function | Optimizer | Epochs |
| 32 | categorical cross entropy | Adam | 10 |

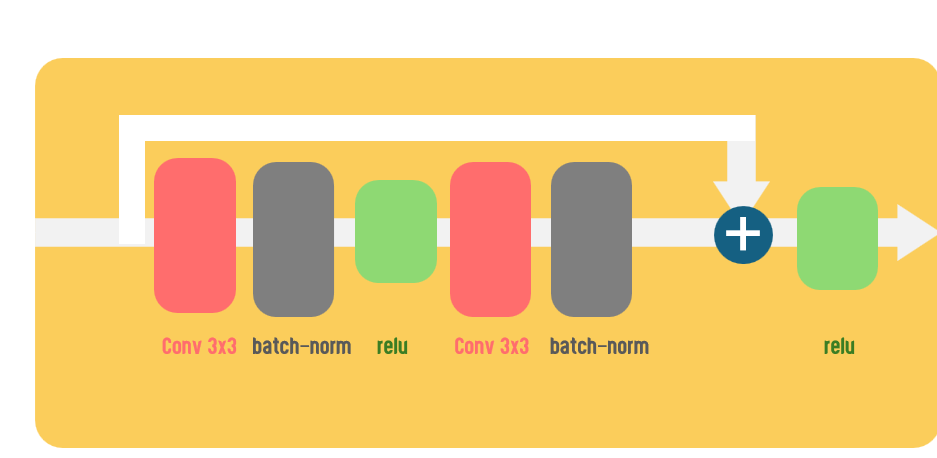
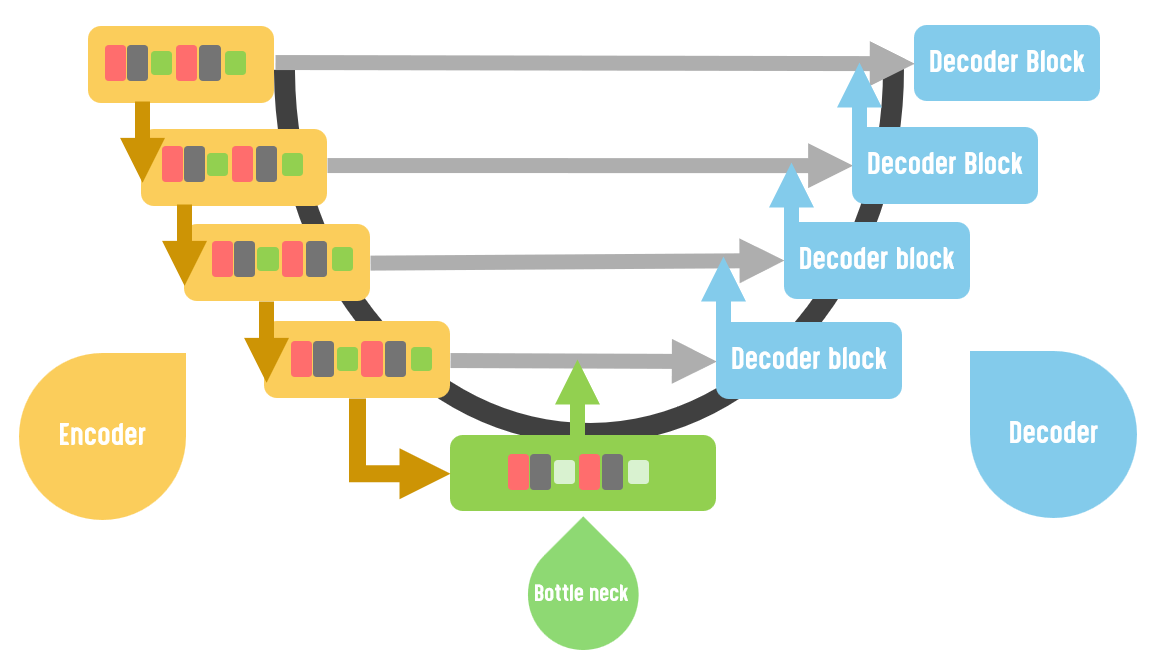
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | Accuracy | Precision | Recall | F1-Score |
| Train (%) | 99.432 | 100.00 | 99.430 | 99.437 |
| Test (%) | 98.74 | 100.00 | 98.26 | 98.36 |

Performance in the first approach

* **Approach Two:**

Second approach based on giving the model the internal part of the lung isolated. For lung isolation from the CT scan, we made segmentation model generate mask for lung.

**Segmentation model architecture**:



architecture based on U-net & its encoder ResNet trained on CT scan data and lung masks

|  |  |  |  |
| --- | --- | --- | --- |
| Batch size | Loss function | Optimizer | Epochs |
| 32 | Custom | Adam | 100 |

Custom loss = class weights x categoric cross entropy + dice loss

(i.e.) there was unbalancing in background and lung part in masks so we calculate weights which will be multiplied in the classes for balancing

Before and after adding class weights to the loss function

|  |  |  |
| --- | --- | --- |
| metrics | IOU | F1-Score |
| before | 0.8753% | 0.9335 % |
| after | 0.9484% | 0.9735% |

A group of colorful squares

Description automatically generated

Classification model architecture is the same but we trained it on CT scan masks applied to them to isolate the lung

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | Accuracy | Precision | Recall | F1-Score |
| Train (%) | 99.714 | 100.00 | 99.711 | 99.715 |
| Test (%) | 95.675 | 94.273 | 95.675 | 95.579 |

Performance in the second approach