Date: 25/Jan/2022

NODE-21 Challenge Submission

(Nodule Detection on Chest x-rays)

Methodology:

The objective of this work is to detect nodules on chest x-ray images. The nodule present in chest x-rays are very tiny objects and of low contrast. The proposed algorithm is based on the YOLOV4 (you look only once – version 4) model because of its excellent accuracy on small object detection. YOLOV4 is a single-stage object detector to perform detection and classification in a single task using convolutional neural networks. In this experiment, we used initial weights trained on the standard COCO dataset and adapt it to re-train the model on the chest x-ray dataset.

The detailed overview of the model architecture and its training strategy is discussed below:

1. Architecture

The model is comprised of three sections:

- Backbone network
- Neck
- Detection head.

The Backbone network is primarily network used as a feature extractor. The neck is a second stage where extracted features are combined and fed to the third stage network. The detection head is responsible to perform prediction that includes classification and regression of bounding boxes.

Fig. 1 explains an overview of architecture.

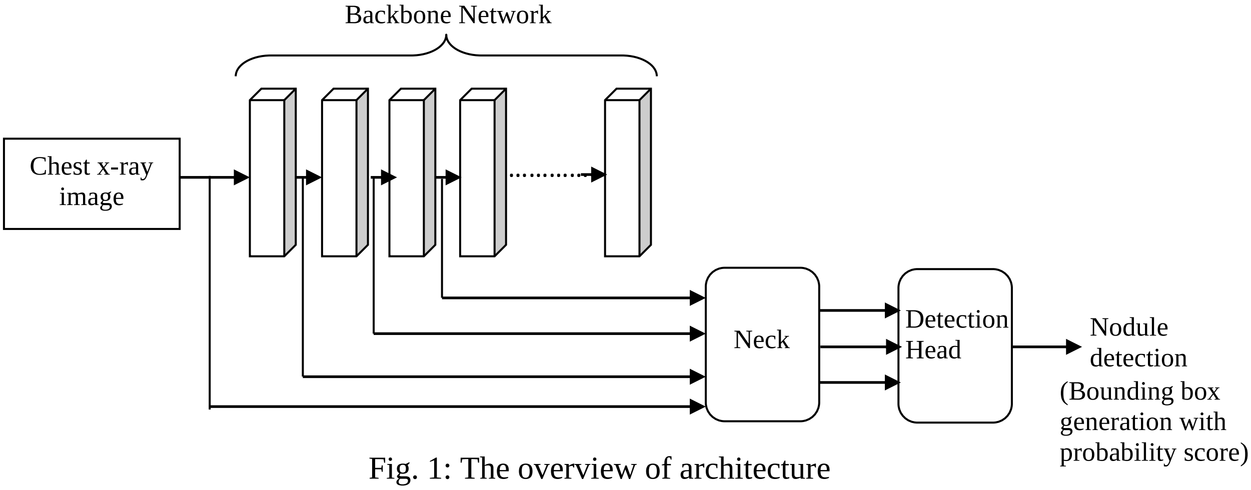

Fig. 1: The overview of architecture

1.1.The Backbone Network:

In this model, CSP-Darknet53 (cross stage partial Darknet53) is used as a backbone network to detect multiple objects of different sizes in a single image. The CSP-darknet architecture is derived from the DenseNet architecture which uses the previous input and concatenates it with the current input before moving into the dense layer. It works on the CSPNet strategy of dividing Dense Block consisting of feature map in two halves and then merging them via cross-stage hierarchy. This strategy helps to decrease the computational complexity.

This is a network with a higher input resolution (608×608), a large receptive field of 725×725, larger number of convolutional layers of size 3×3 with the large number of parameters. Higher input resolution helps in the detection of small objects, and a large receptive field helps to observe the entire objects presents in the image and understand the contexts around the objects. Fig. 2 shows the backbone network.

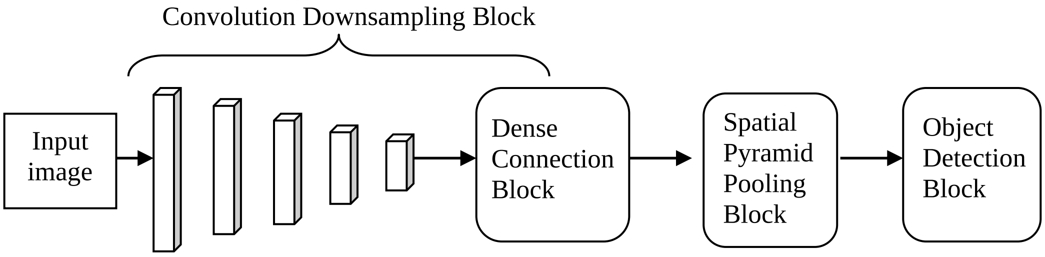

Fig. 2: The Backbone network

1.2.Neck:

This is the second stage that helps to increase the receptive field in the network and select the most significant context feature maps. The essential role of the neck is to collect feature maps from different stages. Usually, a neck is composed of several bottom-up paths and several top-down paths. All the feature maps extracted from the backbone network are gathered in this stage and selected significant features are fed to the head for nodule detection. Neck comprised of many bottom up and top down aggregation paths.

The neck section separately introduces the SPP and PANet as enhanced feature extraction networks that select multi-scale features. The SPP (spatial pyramid pooling) block is added as neck section over the backbone network to increase the receptive field. The SPP block is shown in Fig. 3. SPP is responsible to create fixed size features regardless of feature map sizes. Additionally, it makes the network robust in the case of object deformations. For parameter aggregation, PANet (Path aggregation network) is used for different detector levels. PANet fuses the information from all layers first using element-wise max operation

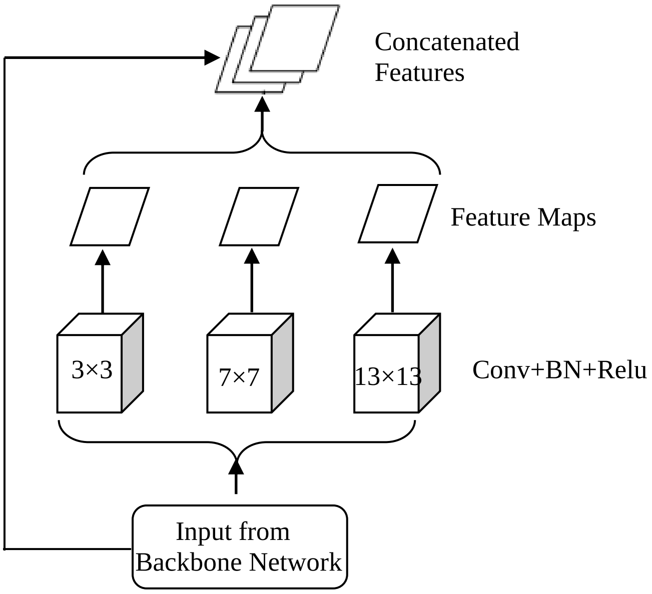

Fig. 3: The Spatial Pyramid Pooling Block

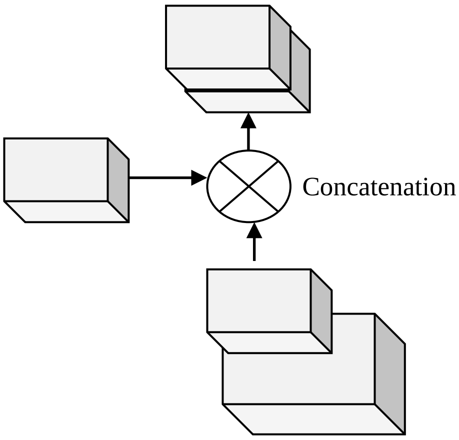

Fig. 4: The path Aggregation Network

1.3.Detection Head:

This stage processes the significant features from the Neck section to predict the bounding box, and their classification score. The detection head is responsible to

- Localization of bounding box.
- Classify what's inside each box.

The detection head follows the principle of one stage anchor based object detection. This section is responsible to perform the final prediction that includes classification and regression of bounding boxes. It provides the coordinate of bounding boxes i.e. x-coordinate and y-coordinate of centre, height, width (x_{centre}, y_{centre}, h, w), and the score of prediction with the label. The detection head is applied to every anchor box.

2. Training Strategy/ Hyper-parameters

In this experiment, we used initial weights of YOLOV4 trained on the standard COCO dataset and adapt it to re-train the model on the chest x-ray dataset. As the training data is not very large so training the model from scratch is not a good idea. Therefore transfer learning is a better way to converge the model faster and to get good accuracy. The network is re-trained for only one class i.e. nodule.

The training hyper-parameters are as follows:
Batch: 64
Subdivision: 16
Maximum batches: 6000
Filters: 18