Task 2: Layout Analysis on structured historical document images

Prateek Keserwani, Shrey Singh, Sanyam Agrawal, Rahul Singhai

PARIMAL Lab, Indian Institute of Technology Roorkee

Deep networks solve task 2 in the sequence of three steps pre-processing, model training and test phase.

1 Pre-Processing Phase

1. Rescaled the raw image(RI) to Scaled image (SI) maintaining a constant aspect ratio of 512 X 512.

2. Decoded the Ground truth(GT) into 3 categories i.e text boundary(tb), background(bg) and text(temp).The text boundary is achieved by applying dilation operation with 5 X 5 kernel and iteration on the text.

2 Model Training Phase

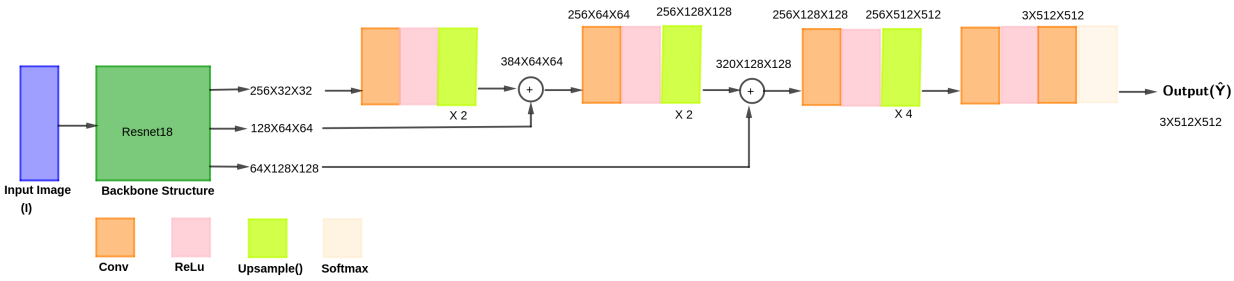

Figure 1: Architecture of Model

1. The SI is passed as an input to the architecture as in Figure 1 which gives an output($\hat{y}$) 3 X 512 X 512.

2. Calculate Dice loss between ground_truth(tb, bg, temp) and predicted output($\hat{y}$)

3. Optimize the loss using ADAM optimizer(lr= 0.001) and $Step\_lr$ scheduler(step_size= 10000, gamma= 0.1).

3 Model Testing Phase

1. Forward passed the testing images in the model with trained weights.

2. Encoded the output obtained back to get the results in the desired format.