CV2: Project Report Template

Anirban Bhowmick

August 2021

1 Introduction

Saliency detection is a fundamental topic in computer vision. It can to solve problems like face detection, object detection, object discovery which can have numerous applications in daily life like in mobile phones camera software to traffic cameras. The main task of the project was to build a model, prepare data sets, train the model and predict the salient regions of the original images. The final salient image predictions are generated as a gray scale image. For this saliency system was developed using deep neural network architecture utilizing the encoder-decoder structure. The model, data set preparation and training was done using Pytorch library and python being the programming language. The details of the experiment are as follows:

2 System Description

For this experiment we use a deep learning network based on encoder-decoder architecture. The architecture has been adapted from the course exercise-5. The encoder part has convolution layers with pooling $F.maxpool2d$ layers in between to down sample the feature maps. The low dimension vector generated by the encoder is fed to the decode which again has convolution layers and up samples $nn.upsample$ the feature maps. Finally during prediction a sigmoid function is used to generate the output image. To implement transfer learning, the encoder part of the model is loaded with the pre-trained weights from VGG16 network [2]. So the encoder parameters are prevented from getting trained. So only the decoder parameters are trained from scratch.

3 Evaluation and Experimental Details

The Dataset provided was already distributed into train, test and valid data sets. The Train and valid data-set had both images and salient fixations(ground truths) while the test set had only the images for testing. The training set had 3006 colour images(dimensions 3x224x224) and 3006 grayscale fixation images(dimensions 1x224x224). The validation set 1128 colour images(dimensions

3x224x224) and 1128 grayscale fixation images(dimensions 1x224x224). Finally the test set had 1032 colour images(dimensions 3x224x224). The loss function used here is Binary cross-entropy loss. The optimizer is Adam optimizer[2] with learning rate 0.001. The experiment was run for 30 epochs with each batches of 16. In the beginning the experiment was also performed with SGD optimizer but it did not yield satisfactory results. So after a lot trial and error Adam optimizer[1] was used. The training loss at the end of 40 epochs stands at 0.16 and the validation loss at 0.71.

4 Results

The model was trained for 30 epochs with batch size 16. The training and validation loss curves are shown below in.

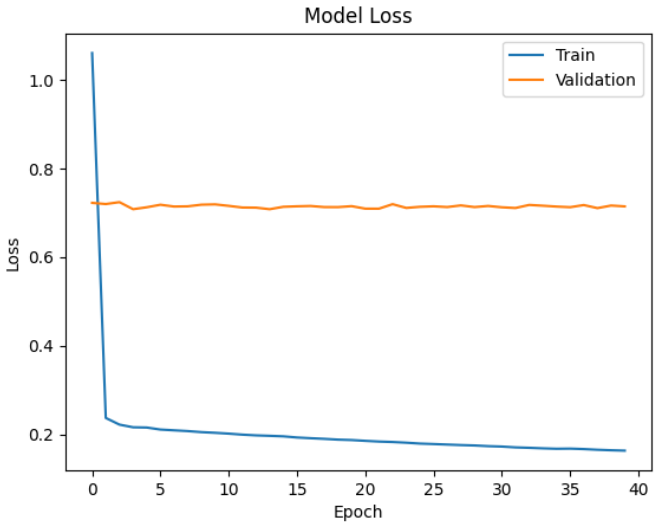


Figure 1: The train-validation loss curve.The training loss stabilizing around 40 epochs.

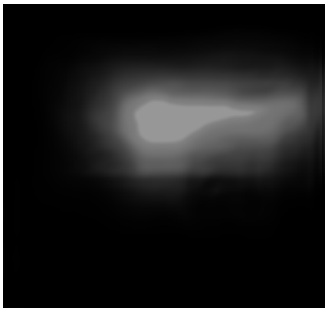
Some of the predicted samples and their respective target values are given below.

(a) Image

(b) Predicted Salient image

Figure 2

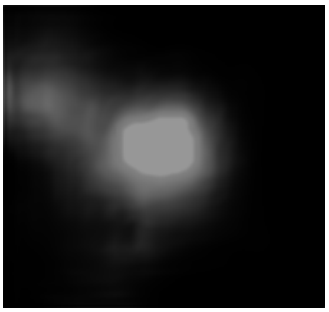
(a) Image



(b) Predicted Salient image

Figure 3

(a) Image



(b) Predicted Salient image

Figure 4

5 Conclusion

Although the saliency predictions were not perfect it could be further improved by training more images. However this project provided a good hands-on on

this area.

References

[1] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.