Deep Learning Project

Comprehensive Report

Topic-

Comparison and Analysis of different datasets for Salient Object Detection with focus on UNet and U2Net

By

Abichal Ghosh **Ninad Mhalgi**

2018A7TS0172G **2018A7TS0162G**

Table of Contents

Introduction

In this work, the authors have focused on 2 deep segmentation models, namely UNet [1] and U^{2}Net [2]. UNet is a simple segmentation model while U^{2}Net is a much more complex model which uses individual UNet structures as its building blocks.

Hence one of the motivations of the work is to understand how well the complex U^{2}Net structure is able to perform in comparison to its building blocks. This has been done by training both UNet and U^{2}Net and then testing both the trained models on multiple test datasets.

The results of the models on the multiple datasets have then been compared both between the 2 models and among itself in order to understand the following -

a. How UNet performs on various datasets
b. How U^{2}Net performs on various datasets
c. How U^{2}Net performs in comparison to its building block the UNet

This comparison gives insights on the complexity of the 2 models. Finally, after rigorous experiments, the best

Datasets

In this work, multiple datasets have been used for training and testing both the UNet and the U2Net models. The datasets which have been used in this study are the following -

1. MSRA Select
2. DUTS
3. ECSSD

MSRA Select

The MSRA Select Dataset is a dataset which has been modified from the MSRA 10K dataset. The MSRA dataset has a total of 10000 images. The MSRA 10K dataset contains over 10k images and their corresponding masks.

- MSRA 10K Dataset for SOD is the dataset that we used for training UNet and 500 images from this was used for training U2Net.

Balanced? : The MSRA 10K dataset is definitely balanced in terms of variety of images ranging from cityscapes, humans, animals, scenes and many more. Further, the MSRA 10K has varying objects of varying sizes, in varying locations meaning there is almost no class imbalance, object scale imbalance and object location imbalance respectively. Hence, this dataset is well balanced.

Relevance : This is a well known dataset for SOD Training due to the variety of images, salient objects and accurate masks. Due to this reason, this dataset is relevant to the paper.

Limitations : Quite a few images in MSRA 10K mostly contain only one salient object. Also most of the images have clear salient objects with simple backgrounds. A better and more realistic dataset would be one with more realistic scenes with non salient objects.

DUTS

DUTS is the primary dataset which has been used in the paper.

Balanced? : The DUTS dataset is definitely balanced due to a variety of images ranging from cityscapes, humans, animals, scenes and many more. Further, due to the vast amount of images.

Relevance : As this dataset was used by U2Net paper itself, the relevance is justified.

Limitations : In DUTS, very few images are present which have a good resolution in comparison to ECSSD and MSRA.

ECSSD

- ECSSD Dataset for SOD is the dataset that we used for training U2Net

ECSSD has more natural images in its dataset. It has 1000 images which are semantically meaningful but structurally complex images for evaluation.

UNet

UNet is a deep image segmentation model which has a U-shaped structure.

The UNet architecture builds upon fully convolutional networks. The architecture has a structure of an encoder-decoder like structure with a contracting path corresponding to the encoder and an expanding path corresponding to the decoder. The contracting path aims at capturing context and the expanding path aims at localization.

The UNet model has been trained on the MSRA Select Dataset for 15 training epochs and with a learning rate of 0.000065.

U^2Net

U^2Net is a deep salient object detection model, which uses a 2-level nested U-architecture. The inner U-structure is similar to the architecture used in UNet. Hence, U^2Net can be approximately visualized as a Unet with the building blocks as UNet themselves, as apparent from the structure of the model.

The U^2Net model has been trained on both the MSRA Select and the ECSSD Select Datasets for 15 training epochs and with a learning rate of 0.00065. Various epochs and learning rates were considered for this, but these choices of hyperparameters were the closest to optimal. In the further models, the same hyperparameters have been used in order to aid with comparative study of the performance of the models

Experiments

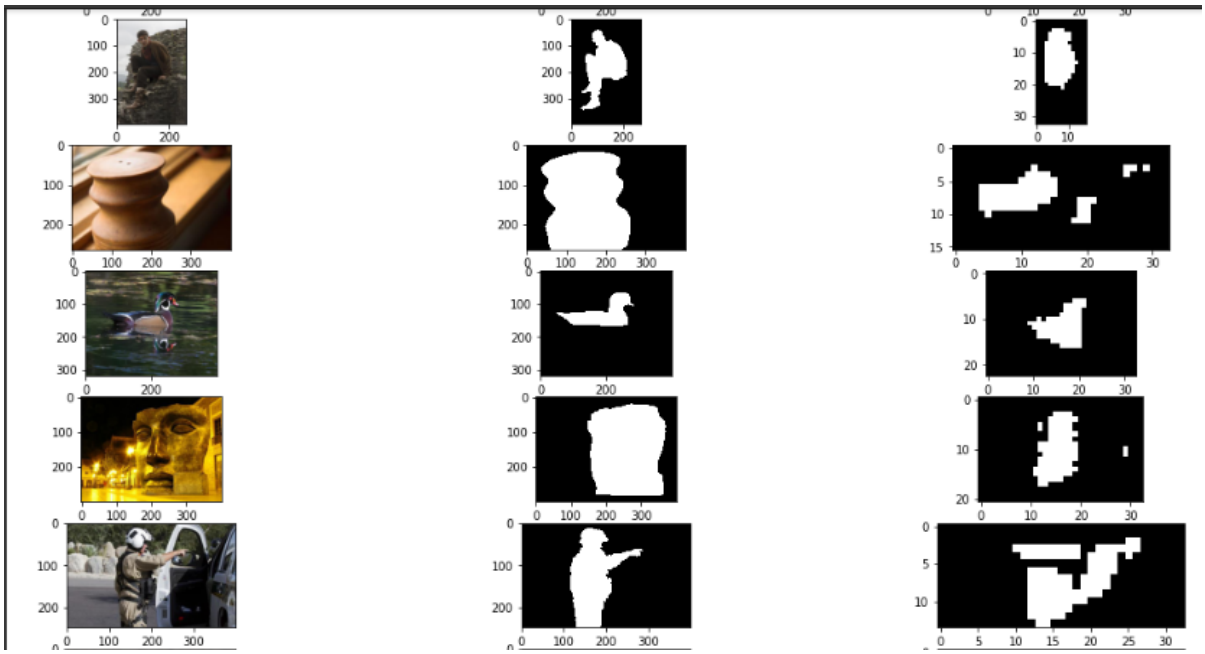
The trained UNet and U^{2}Net models have been tested on different test/validation datasets. The validation results of these models have then been used for 3 types of comparisons which are-

1. Understanding UNet
2. Comparison of U^{2}Net performance on different datasets

Understanding UNet

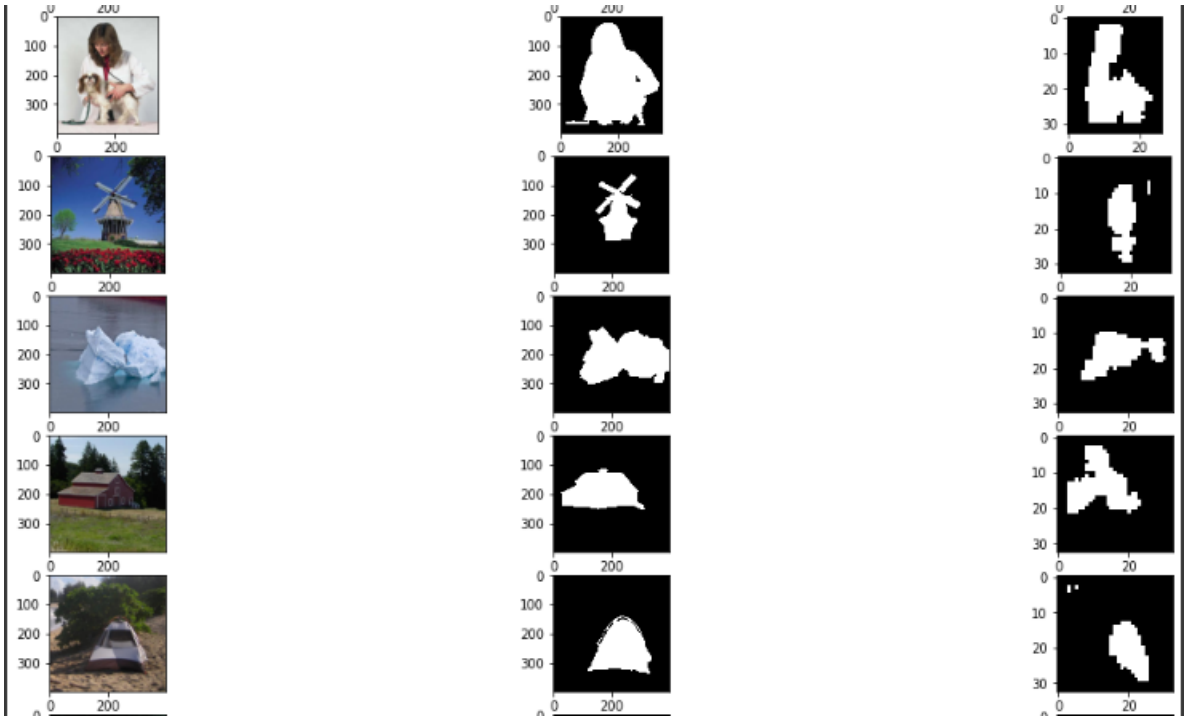
In order to understand how the trained UNet model runs on unseen data, the model has been trained on the MSRA select dataset.

Following this, the model has been validated on MSRA 10K select test, DUTS test and ECSSD test datasets with the results obtained being the following-



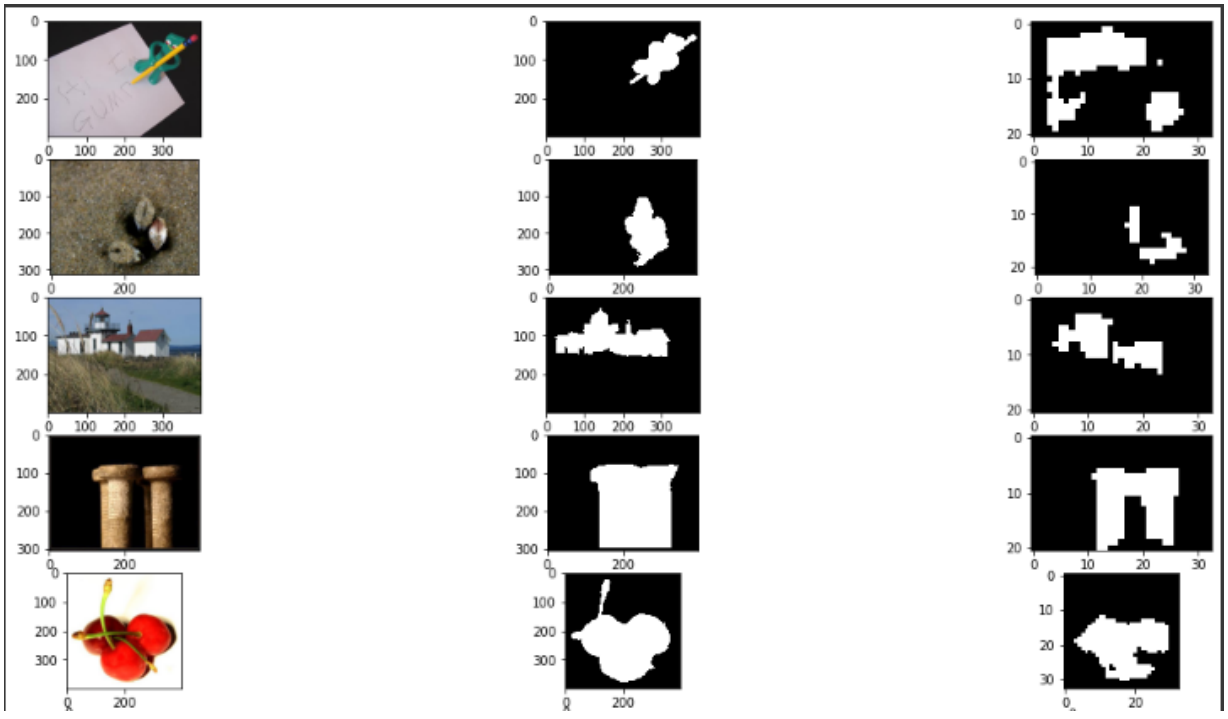
(a)Image (b) mask (c)predicted

Fig : UNet testing on ECSSD validation set.



(a)Image (b) mask (c)predicted

Fig : UNet testing on DUTS validation set.



(a)Image (b) mask (c)predicted

Fig : UNet testing on MSRA validation set.

With this, it is seen that the validation masks predicted by Unet are extremely rough and are not good enough approximations. Hence, further experiments have only been done with U2net trained on MSRA Select and ECSSD datasets.

Comparison of U^{2}Net performance on different datasets

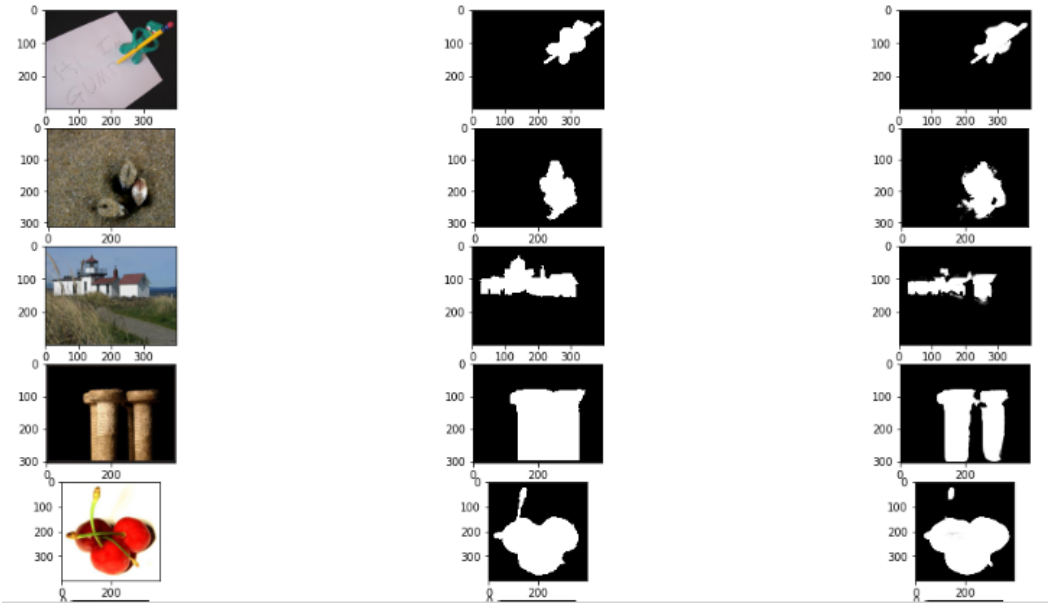
For this comparison, U^{2}Net has been trained on two different datasets which are MSRA and ECSSD respectively. Each of these models have then been tested on all the three datasets.

Since the DUTS test dataset has types of images which are completely different from the ones found in MSRA and ECSSD training dataset, it can be said that the validation performance of these 2 models on the DUTS dataset gives us a correct measure of how well each of the model performs on unseen data.

The results of validation are as follows-

Validation on MSRA

- Model trained on MSRA(similar to seen data)



(a) Image (b) mask (c)predicted

Fig : U^{2}Net(MSRA) testing on MSRA validation set.

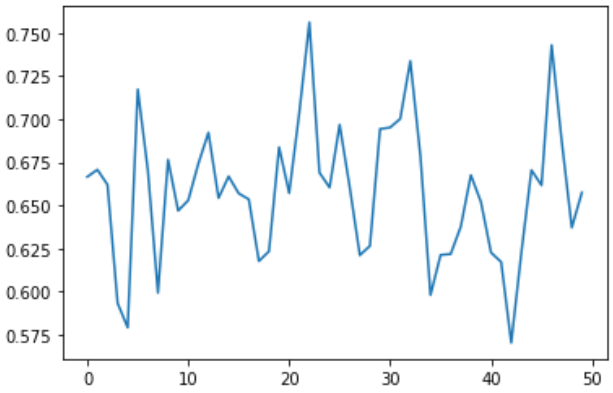
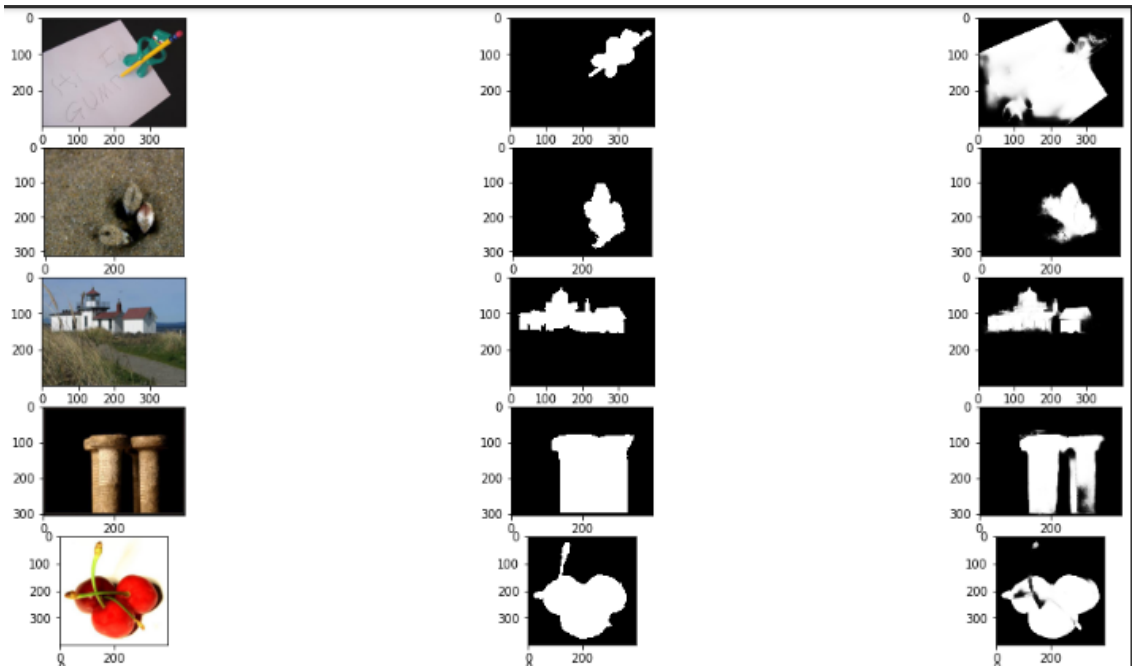


Fig: Validation loss graph for 50 images in validation dataset of MSRA.

Average validation loss is 0.6579421103000641.

- Model trained on ECSSD(unseen)



(a)Image (b) mask (c)predicted

Fig : U^{2}Net (ECSSD) testing on MSRA validation set.

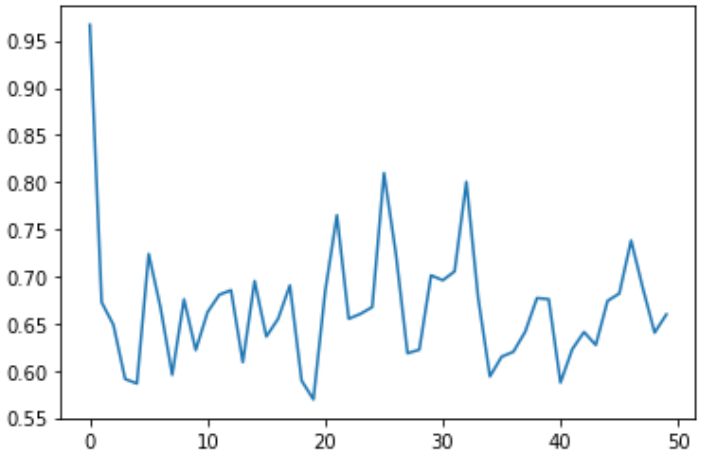
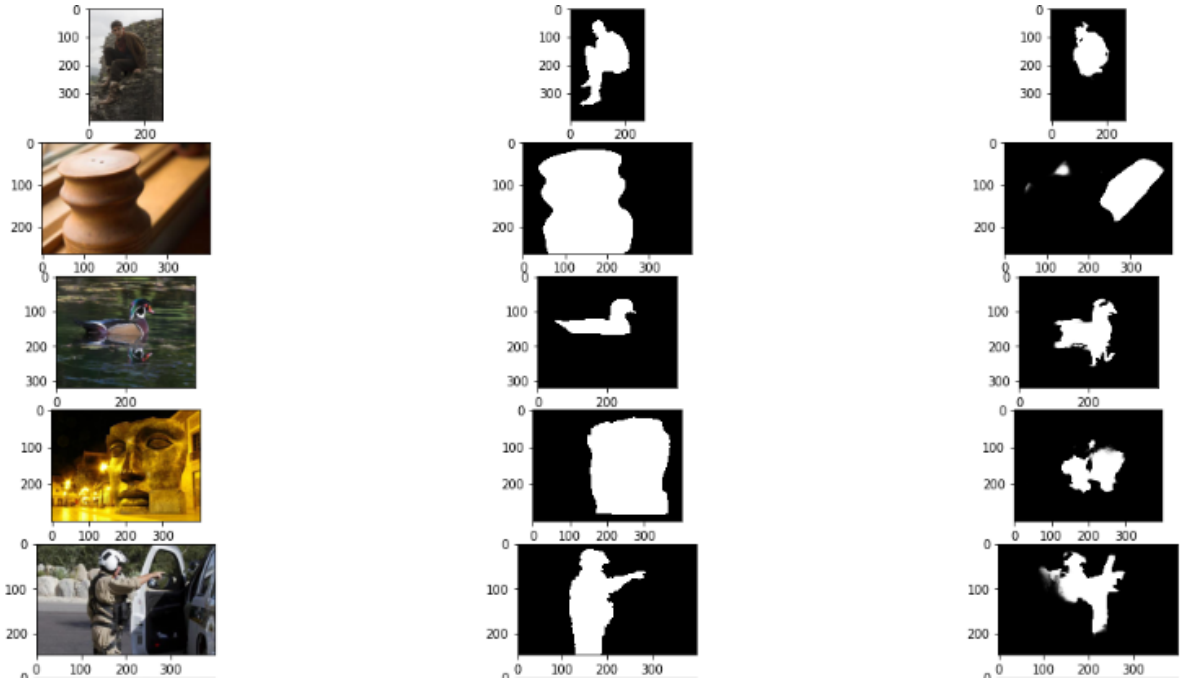


Fig: Validation loss graph for images in validation dataset of MSRA.

Average validation loss is 0.6677632081508637.

Validation on ECSSD

- Model trained on MSRA(unseen)



(a)Image (b) mask (c)predicted

Fig : U^{2}Net (MSRA) testing on ECSSD validation set.

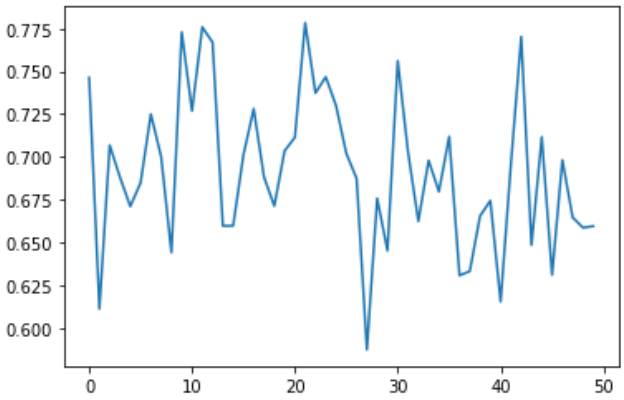
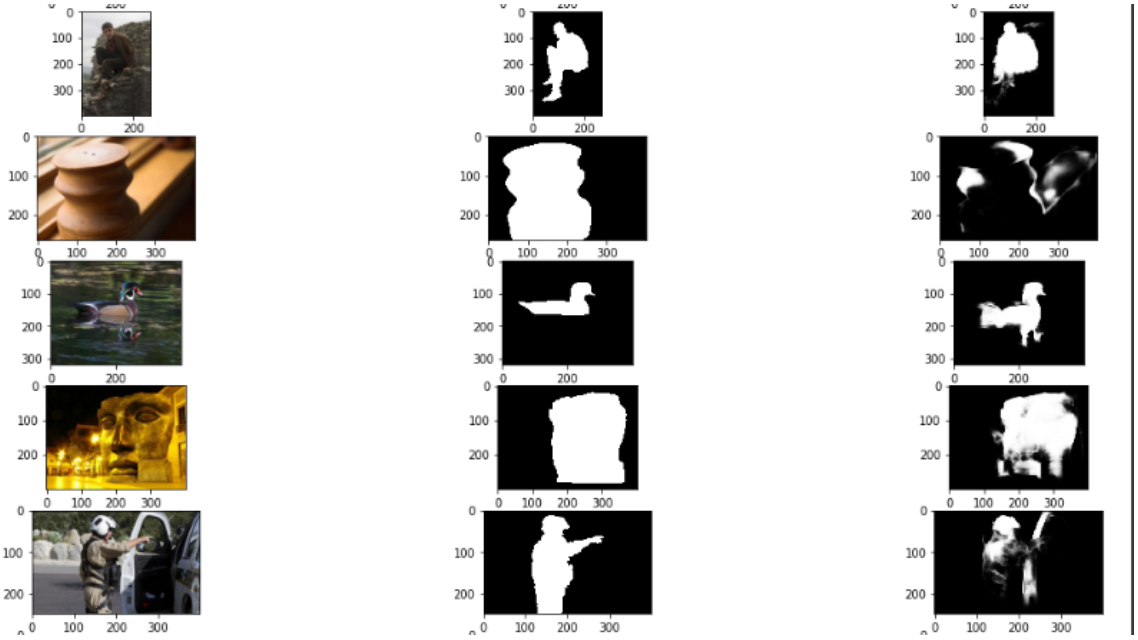


Fig: Validation loss graph for images in validation dataset of ECSSD.

Average validation loss is 0.6920081424713135.

- Model trained on ECSSD(similar to seen data)



(a)Image (b) mask (c)predicted

Fig : U^{2}Net (ECSSD) testing on ECSSD validation set.

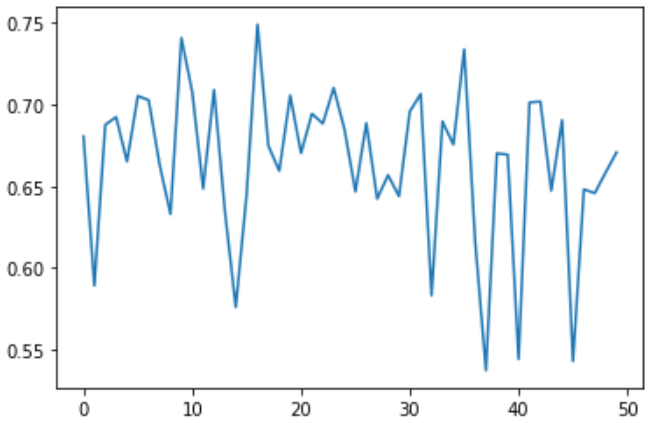
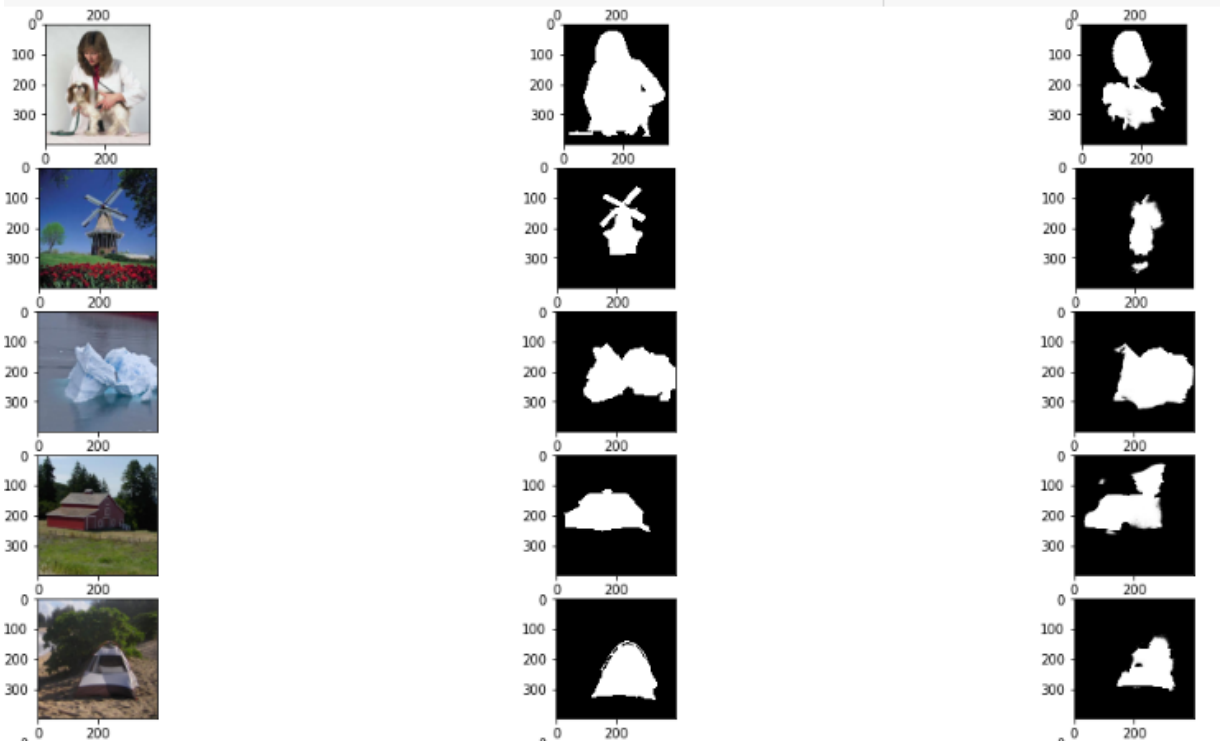


Fig: Validation loss graph for images in validation dataset of ECSSD

Average validation loss is 0.6644891130924225.

Validation on DUTS

- Model trained on MSRA(unseen)



(a)Image (b) mask (c)predicted

Fig : U^{2}Net (MSRA) testing on DUTS validation set.

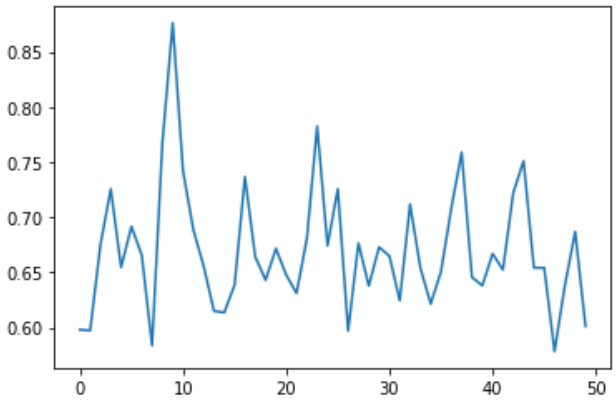
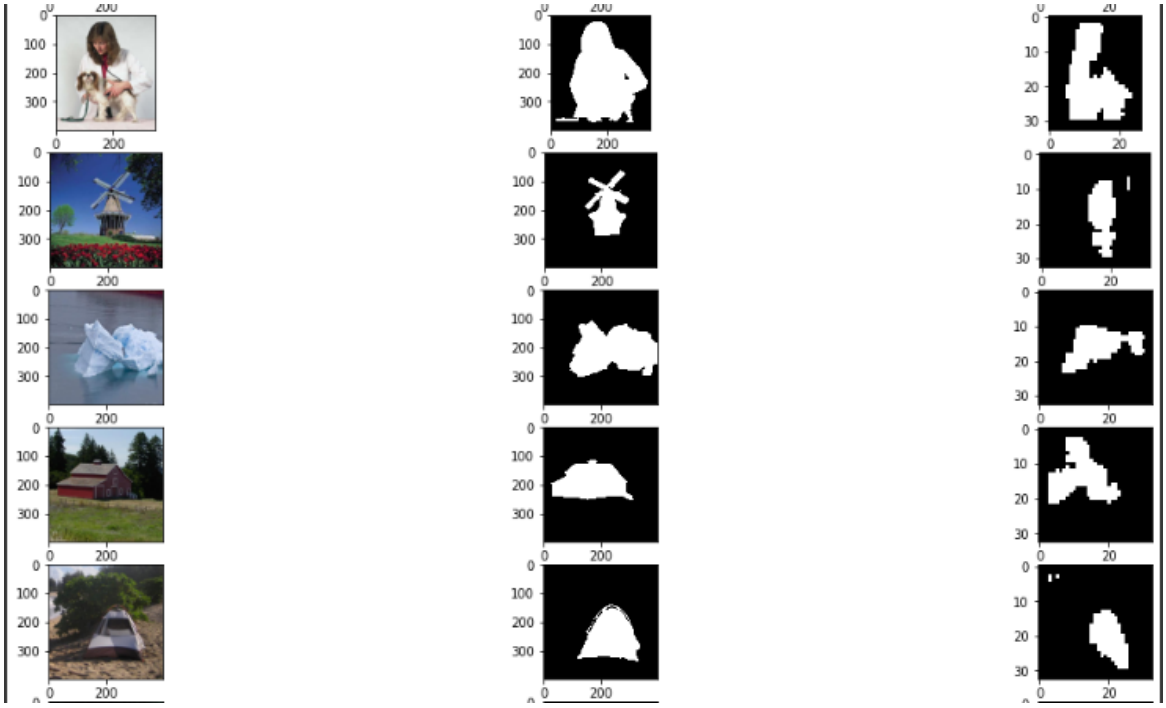


Fig: Validation loss graph for images in validation dataset of DUTS

Average validation loss is 0.69179400253668719

- Model trained on ECSSD(unseen)



(a)Image (b) mask (c)predicted

Fig : U^{2}Net (ECSSD) testing on DUTS validation set.

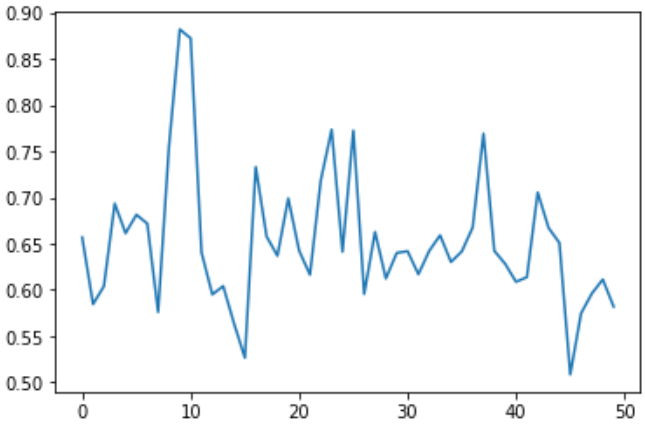


Fig: Validation loss graph for images in validation dataset of DUTS

Average validation loss is 0.6702016389369965.

Conclusion

` From the results of the above experiments, the following can be concluded.

FIrstly, as is obvious, each of the trained models perform significantly better when they are tested against the dataset corresponding to which they had been trained with. For example, Model trained with MSRA when tested with MSRA test dataset, the validation loss is 0.657 which is lower than the validation loss of 0.667 of the model trained with ECSSD and tested on MSRA test dataset. Similarly, 0.664 validation loss corresponding to model trained with ECSSD and tested on ECSSD test dataset is lower than the 0.69 validation loss of model trained with MRSA and tested on ECSSD test dataset.

The final conclusion is observed when we compare the validation losses of the 2 models on unseen data. The first instance of this is when the validation loss of the model trained on ECSSD when tested on MSRA test dataset is 0.667 which is lower in comparison to the validation loss of 0.692 of the model trained on MSRA and tested on DUTS test dataset.

A similar conclusion is found when we compare the validation losses of the 2 models again on the unseen data of the DUTS dataset(similar to both the models being tested on unseen data). Here, it is seen that the validation loss of the model trained on ECSSD when tested on the DUTS test dataset is 0.67 which is lower in comparison to the validation loss of 0.691 of the model trained on MSRA and tested on DUTS test dataset.

This leads to the conclusion that U2Net models trained on ECSSD are able to predict desirable features even in images where the objects to be detected are camouflaged, blended with surroundings, have low contrast difference with the surroundings, or are difficult to detect. From this, it can be concluded that the ECSSD Dataset is a much better dataset for salient object detection in comparison to the MSRA dataset. However, the only possible downside is that the number of images in ECSSD is much lower than that present in MSRA.