Skin Lesion Segmentation using SegNet-U-Net Ensemble

Prashant Brahmbhatt[a], Rishi Chaturvedi[a], Siddhi Nath Rajan[a]* Bernd Markscheffel[b]
[a]*Department of Information Technology,*
IMS Engineering College, Ghaziabad, India.
[b]*Technische Universität Ilmenau (Germany)*
*sn.rajan@imsec.ac.in

Received: *05.12.2019,* ***Accepted:*** *16.12.2019*

ABSTRACT

The paper is all about applying an ensemble method for the segmentation of the well-known problem of Skin Lesion. The paper combines the conventional approach with the ensemble principle to achieve reasonable performance. The secondary aim is to reduce the effort of pre and post processing of the images. The dataset used for the training and testing of the approach is the PH2 dataset which is composed of the dermoscopic images of skin lesions and their respective ground truths which are obtained by the conventional manual method.

Keywords - Skin Lesion Segmentation, Dermoscopic, Deep Convolutional Network

1. INTRODUCTION

The American Cancer Society provides the stats which indicate that melanoma cancers are 1% of all the diagnosed ones. Although the number is less but the death proportion in melanoma is very high. Also, a significant number of deaths are due to the Non-Melanoma cancer. The predicted death numbers for the year 2019 is 7,230 which includes 4,740 men and 2,490 women from the melanoma cancer alone (Melanoma: Statistics | Cancer.net, 2019). The numbers on a global level with reference to the World Health Organization are 3 million from Non-Melanoma and 1,32,000 from melanoma(WHO, 2019). The skin cancer can be treated provided that it is diagnosed timely and the expected life of the treated patient increases greatly. So, we can infer that how important it is to correctly identify and diagnose the skin lesion. The skin lesions are broadly Melanoma and Melanoma with each one having their sub-categories as well. The lesion images are segmented before they can be classified, the binary masks are to be created to crop out the lesion area from the given image, the cropped images.

Before the automatic segmentation was proposed first, the traditional methods were being used in which the images were being segmented by the expert dermatologists sitting under the microscopes for prolonged periods. This method had quite a few problems associated:the time it took to be performed, the complexities, the dependence on the observer and his level of expertise and the variations in the number of available experts in different cities and countries.

Since the proposal of the automatic segmentation, several different neural networks have been proposed claiming better performance over the previously SOTA network. Some of the approaches are the FCN (Yuan, 2017), VGG (Lopez, 2017), Fast-RCNN (Girshick, 2015) and U-Net (Md Zahangir Alom, 2018).

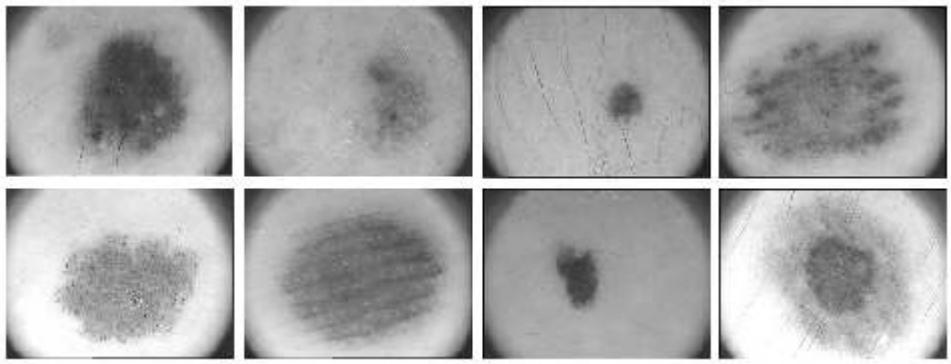


Figure 1: Sample microscopic images of the skin lesions found on the human skin

2. THE MODEL

The first model is the U-Net model. U-Netis a neural network architecture created for the purpose of Bio Medical Image Segmentation (Ronneberger, 2015). There are two paths in the architecture with the first path being the encoder or the contraction path to capture the essence of the image and the second path is the decoder or expanding path whose purpose is to accurately localize the lesion.It can receive image of any size which clarifies the absence of a dense layer in the network.

The second model is SegNet whichis a deep neural network (Kendall, 2017). The basis of SegNet is the encoding-decoding model used for pixel-wise segmentation. It was proposed originally and developed by Computer Vision and Robotics Group Members at the University of Cambridge, UK (Alex Kendall, 2015).Semantic Pixel Wise labelling is the purpose of SegNet, it was initially used for the street images multi-class segmentation. There were twelve different classes that each pixel could belong to as per the data. The SegNet has encoding sequences which are non-linear layers and each layer has a matching decoder. Another classifier is inserted at the end of the network to give the final classification output. Several improvisations on the hyperparameters have been done to increases the accuracy of the original SegNet results.

2.1 NETWORK ARCHITECTURE

The architecture of the network are the ones already proposed, the traditional U-Net-224 architecture which is originally designed to received 224 x 224 images has been modified to take resized input images suitable for our scenario. The SegNet architecture is the one proposed with Binary Cross-Entropy as loss function and two dense layers different to the conventional architecture(Brahmbhatt, 2019).

2.2 BAGGING IN ENSEMBLE

Ensemble in machine learning is used to get optimal results by combining the several fitted models for a single problem, it helps in leveraging the correct predictions made by one model and helps in isolating the incorrect ones. There are many techniques that are used for ensembling a model. The one used here is the bagging of models.

In the bagging technique, it combines the principles of bootstrap and aggregation, which justifies the name ***B***ootstrap ***AGG***regat***ING***. In this process the different learners are fitted on the data which are considered as weak learners and then some aggregation method is used to have final output. Generally, an averaging process is used which has also been used here in SegNet– U-Net ensemble.

2.3 LOSS FUNCTION

The loss functions used here are two different mathematical functions for the different networks.In the U-Net U-Netarchitecturethe jaccard distance is used for measuring loss.The jaccard distance is opposite to the jaccard index (intersection over union). Jaccard Index measures the dissimilarity between two sets, it is obtained by subtracting the jaccard index from 1. We can also subtract the union and intersection then divide it by the union to obtain equivalent result. The jaccard distance is in the form below:

$$d_j(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

In the SegNet, the loss function used is the binary cross-entropy. A binary cross-entropy measures how much different the correct values are from the predicted values for each class in the problem and averages out the errors as per the class to get the loss.

Only two classes are present in this problem, either white or black (0 or 1) for every mask image. So, rather than originally proposed categorical cross-entropy the loss function used here is the binary cross-entropy.

The binary cross-entropy is in the below form:

$$L(y,\tilde{y}) = -\frac{1}{N}\sum_{1=0}^{N} (y * \log(\tilde{y}_1) + (1-y) * \log(1-\tilde{y}_1))$$

2.4 IMAGE AUGMENTATION

To enhance the model and increase the robustness the image augmentation has been performed which also reduces the chances of overfitting. It also increases the size of our dataset which in this scenario is even more useful due to the smaller number of images available for training. The techniques used for augmentation are the image rotation(Yuan, 2017) and horizontal flipping(Perez, 2018). In the flipping the images are flipped horizontally resulting in the mirror images of the original images. In the rotation the images are randomly rotated in the range [-40°, $+40^{\circ}$].

The corresponding masks to each of the original images are also exactly transformed which maintains the correct orientation of the ground truths and their feature images.

The augmentation procedure was performed on the 150 images out of the 200 feature images, 50 images are held back for the final test performance measurement. After transformation the training size has increased to 450, out of those 450, 90 images are separated for the validation set which will be used along with the training set to measure performance after each epoch and contribute to better training.

2.5 TRAINING

The training is finally performed over the remaining 360 images in the training set and validated over 90 images of the PH2 dataset(T. Mendonça, 2013).For the U-Net, the total trainable parameters are 31,454,721 and non-trainable parameters are 12,032, while in the SegNet the trainable parameters are 33,377,795 and non-trainable parameters are 15,874. The implementations of both the networks are in Python 3.5 using Tensorflow'sKeras. Kaggle'sIPython notebooks(Anthony Goldbloom, 2010) are used as platform for the entire problem. The total training time for U-Net over 450 images is 12.15 minutes and SegNet took around 18.19 minutes.

2.6 OPTIMIZER

There are two different optimizers used for the different networks, the Adam optimizer and Stochastic Gradient Descent (SGD). In the U-Net the employed optimizer is the Adam optimizer and the learning rate which is an important hyperparameter for training is set to 0.003 for the U-Net which is a value generally used as learning rate. For the SegNet the SGD is used in the training and the learning rate used for the same is set to 0.001.

In the SegNet momentum approach is also used which provides an updating rule of weights from the physical perspective of optimization based on the prediction of the next likely heading direction of the optimizer. The advantage of this approach is that a small change results in large speed hike in the learning process. Analogically, the velocities are stored for all the parameters, and used for making the updates. The value of momentum used in for optimization is 0.9.

2.7 BATCH NORMALIZATION

The batch normalization is the process in which learning of the neural network is increased by performing normalization on the weight values of the hidden layers. The concept behind this normalization is same as in the scenario of data analysis or activation values. In both of the proposednetworksafter every convolution layer a batch normalization layer is present.

3. EXPERIMENTAL DESIGN AND RESULTS

3.1 DATA SOURCE

The dataset used in for the purpose of this research is the PH2 dataset of dermoscopic images which is comprised of 200 feature images and their corresponding ground truth label masks. Every image in the dataset is a three channel RGB image and the initial dimension of each image is 572 x 765. The dataset is a publicly available dataset which is used for experiment and study purposes. The acquired database is from the Dermatology Service of Hospital Pedro Hispano, Matosinhos,Portugal.
For training image dimensions for each sample has been reduced to a size of 192 x 256 and same for the matching label masks. It not only reduces the complexity but also the training time without having a significant impact on the results.

3.2 PERFORMANCE EVALUATION

The generated binary mask images in the output of the network are evaluated on different mathematical measures in comparison with the ground truth lesion masks as provided inthedataset.Theaccuracyismeasuredpixel-wise. The used measures are as below:

- Intersection Over Union:The Jaccard index, also known as Intersection over Union. The Jaccard similarity coefficient is a statistical similarity measure to check the diversity among the sample sets. The IOU gives the similarity among sets and the formula is the size of the intersection over the size of the union of the sets.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

- Dice Coefficient: The Dice score is like precision. It measures the positives as well as it applies penalty to the false positives given by the model.It is more similar to precision than accuracy.

$$Dice = \frac{(2 \times TP)}{(TP+FP) + (TP+FP)}$$

- Precision: Precision is a measure which is more focused towards catching the false positives in the results of the model.

$$Precision = \frac{TP}{(TP+FP)}$$

- Recall: Recall is a measure which is targeted towards the actual or the true positives yielded by the model output. In the scenarios where the cost of the False Negatives is greater than recall is the better metric to choose the best model among the possible ones.

$$Recall = \frac{TP}{(TP+FN)}$$

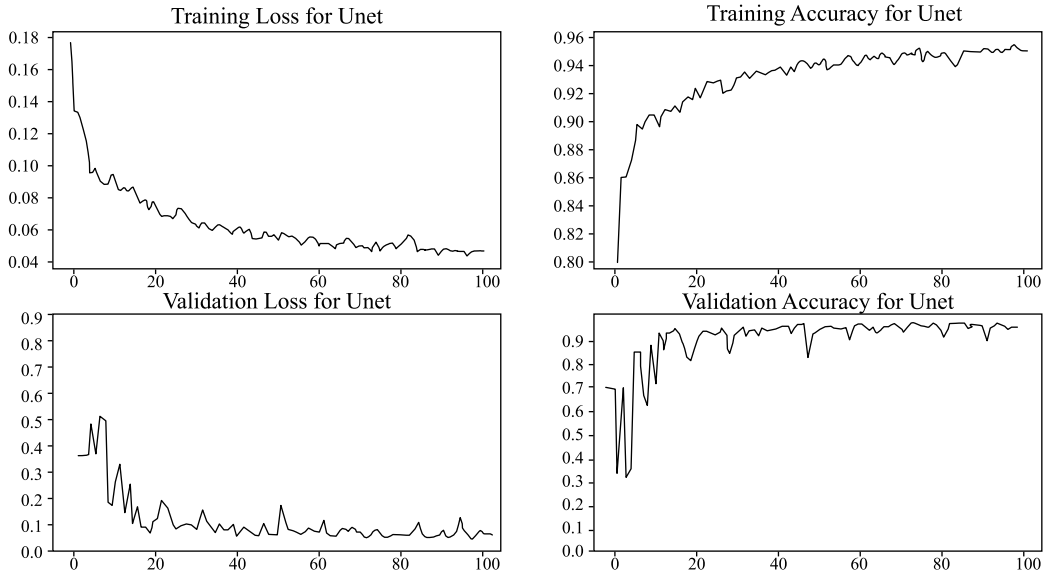
- Accuracy:

$$Accuracy = \frac{TP + TN}{TP+FN+TN+FP}$$

3.3 KEY COMPONENT VALIDATION

The networksproduced the following results after the training procedure, these performance measures do not include the ensemble and are the raw values after the initial training. The training curves of the network corresponding to the training set as well as the validation set are also plotted. The curves include the loss curve and the accuracy curve with respect to the epochs along the horizontal axis. Initially for U-Netthe training set loss is above 0.177 which gradually declined and reached 0.04 and the validation loss began from 0.342 and ended up at 0.0512. The accuracy for the U-Nettraining set increased from 0.798 to 0.954 and that of the validation set from 0.687 to 0.945.

For the SegNet the training set loss is above 0.715 which gradually declined and reached 0.122 and the validation loss began from 0.669 and ended up at 0.161. The accuracy for the SegNet training set increased from 0.605 to 0.975 and that of the validation set from 0.687 to 0.953.



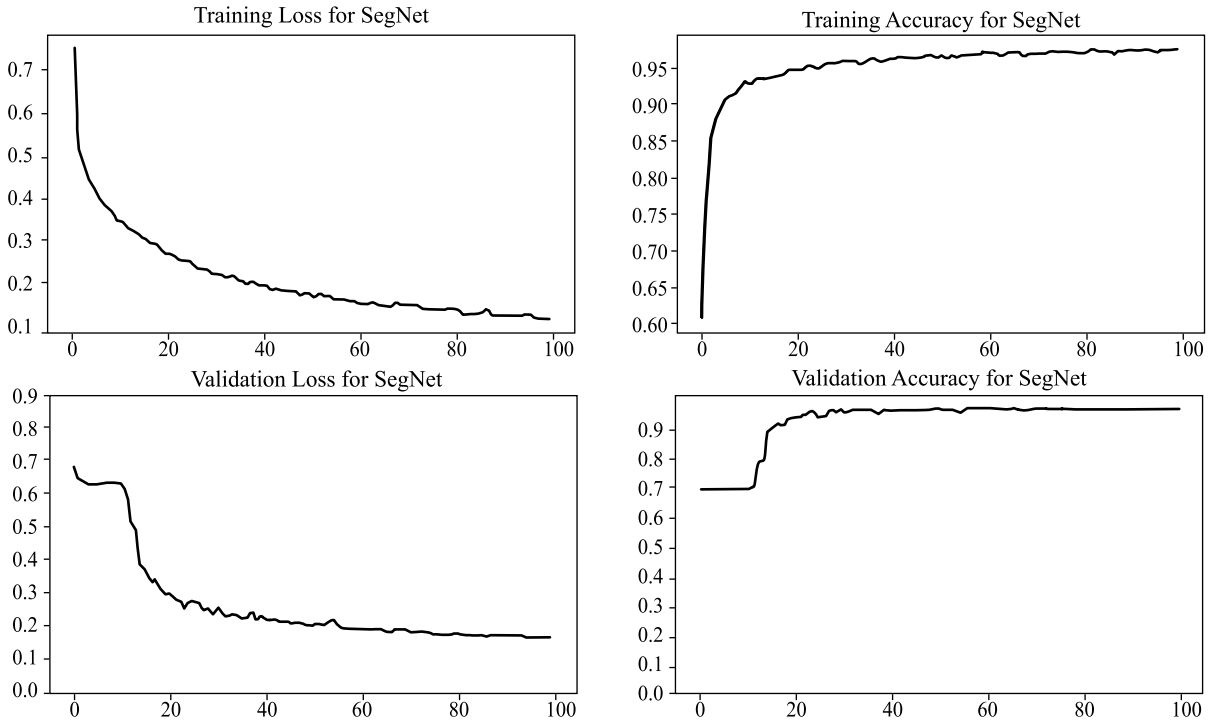


Figure 2: The training and validation curves for U-Net and SegNet.

Table 1: Performance statistics on U-Netafter training for 100 epochs

Dataset	JA	DI	PR	RE	AC	Loss
Test	94.24	89.78	86.86	96.26	94.21	5.76
Validation	94.88	90.12	89.03	94.15	94.50	5.12
Training	95.17	91.00	90.66	94.38	95.06	4.83

Table 2: Performance statistics on SegNet after training for 100 epochs

Dataset	JA	DI	PR	RE	AC	Loss
Test	93.62	80.38	89.24	91.94	94.01	18.79
Validation	94.93	81.86	89.03	93.18	95.37	16.17
Training	97.07	85.24	96.90	96.51	97.94	11.34

3.4 INITIAL PREDICTIONS ON PH2 DATABASE

Now after the training of both the networks we will observe the initial predictions of both the networks on the PH2 database and see how well they are initially performing.

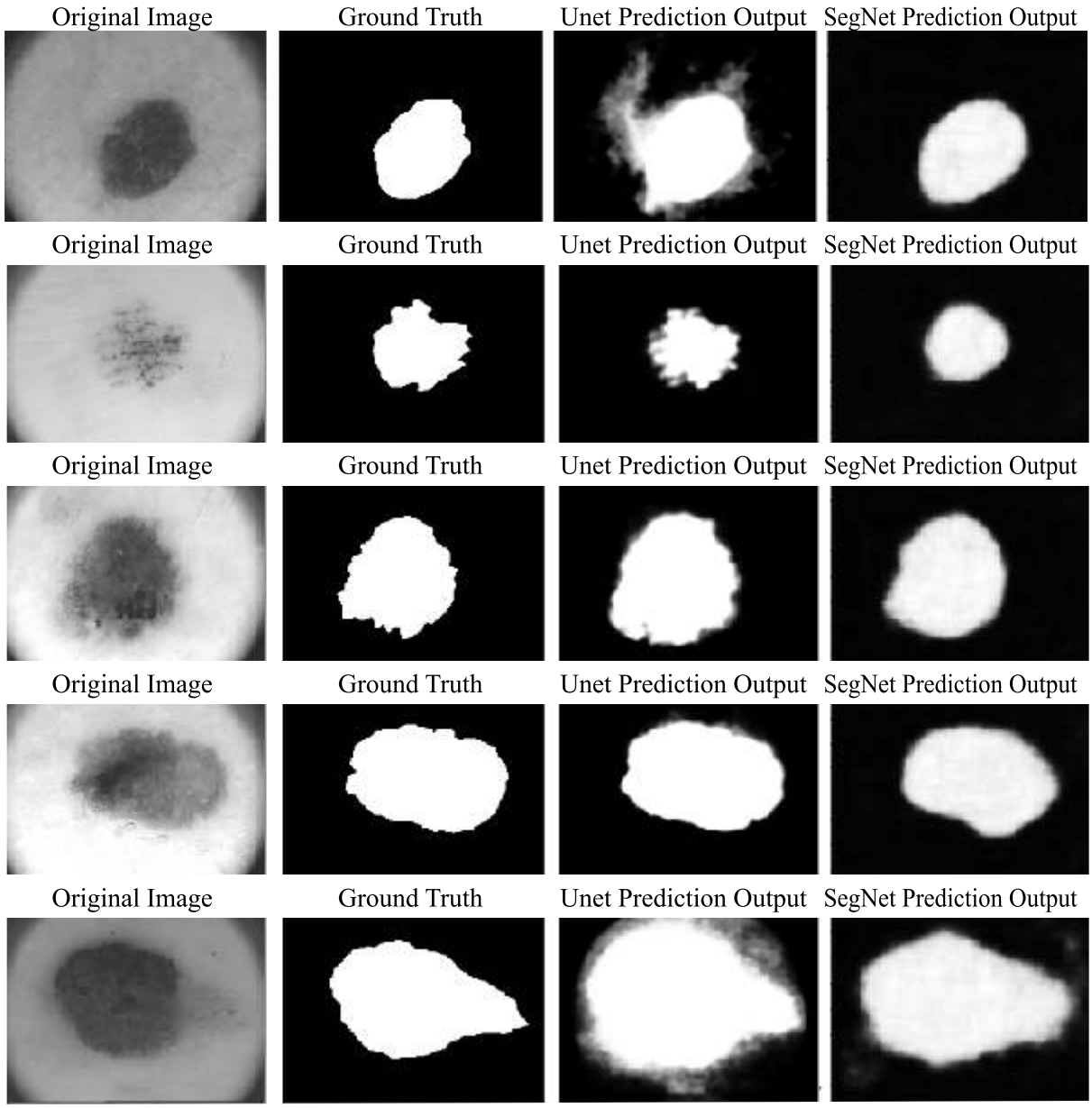


Figure.3: The comparison of the predictions from both networks with original feature images and the ground truths

In the above we can observe that the predictions made by both of the networks are very well, but both are reacting differently to subtle details of the image. Whereas the U-Net is providing more accurate but more noisy predictions, the SegNet is giving vague boundaries but it is less prone to noise compared to U-Net.

So, it is reasonable to combine the outputs of both the networks so that we can balance the features from both the networks and we use bagging to do so. We will average out the values predicted for each pixel from both the networks. We manually do this bagging procedure and sum the predicted values and divide them by 2 to get the final value for each pixel.After the bagging, we use a thresholding technique to remove the gray predictions which are due to the continuous values of the predictions. The thresholding value used is 0.7, any value greater than 0.7 will be rounded-up to 1 and any value less than or equal to it will be 0. The results after the ensemble process and thresholding are as below:

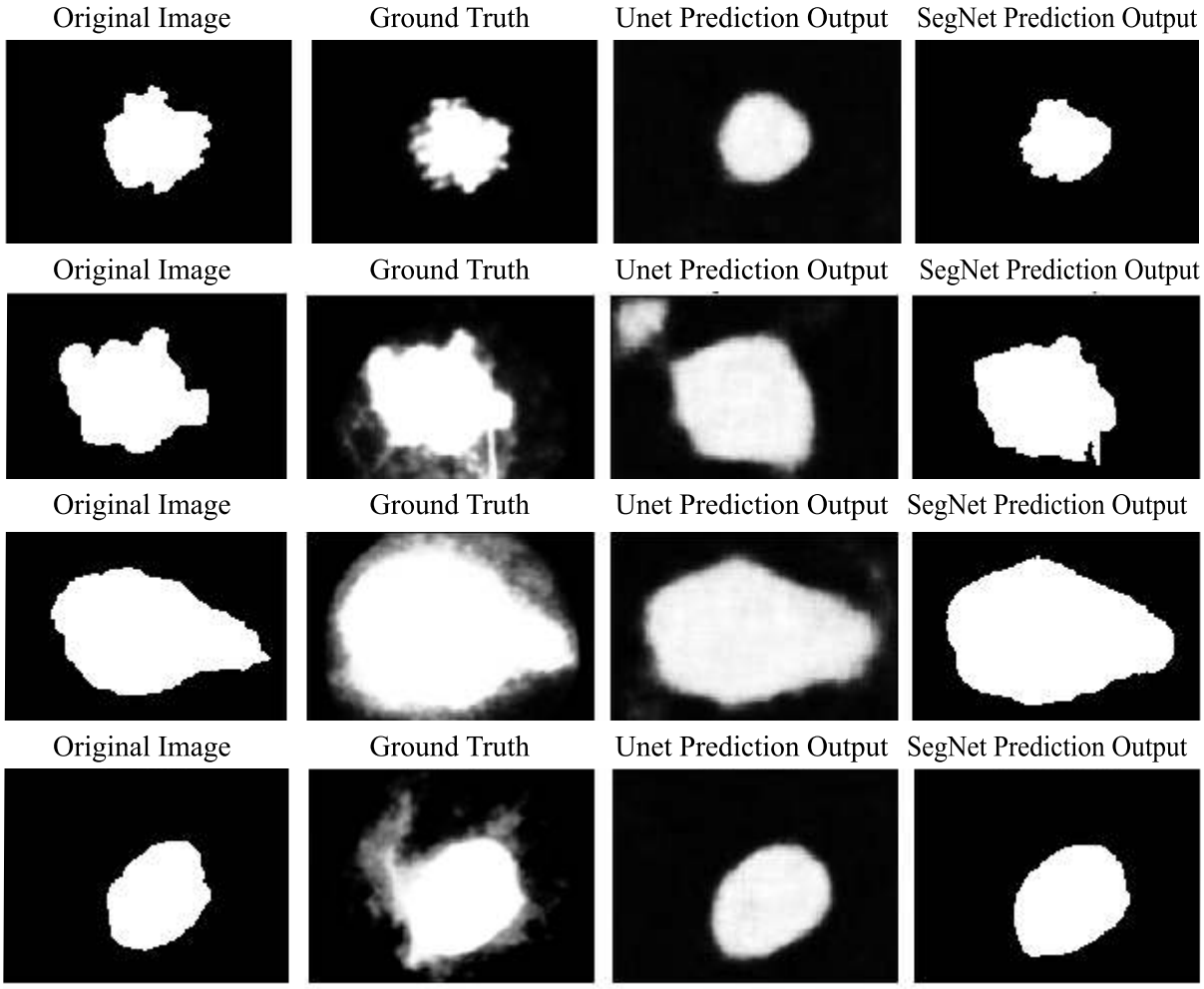


Figure 4: The comparison of ground truths and original predictions with ensemble predictions

We can now see that the final ensemble observations are better than both of the network performances individually. In the ensemble predictions the boundaries are accurate like the U-Net and have less noise like the SegNet which ensures that out bagging approach is providing reasonably better results for each sample.

4. DISCUSSIONS

The proposed method in the paper are only verified on the PH2 dataset. This approach can be extended to the popularISBI 2016 (Gutman, 2016) dataset and ISIC 2017(Codella, 2017) dataset to have an even better idea of how feasible this approach in practical purpose.

There could be more augmentations techniques used like random cropping of images, levelsets (Kenny H. Cha, 2016) and varying brightness and contrasts which could improve the robustness however it would also increase the complexity of the model which already includes training two networks.

The thresholding value that has been used for the discretization of the pixel predictions is merely based on the experimental results and performance and has no theoretical aspect behind choosing that specific value. The approach could also be tested with images size different to that of the one used here, although choosing the larger size of the images will increase the complexity and time but could be need for datasets which have large number of images like the datasets mentioned above which have 2000 images and more.

5. CONCLUSION

In this paper the authors proposed the ensemble of the two networks SegNetand U-Netarchitecture for the problem skin lesion segmentation and successfully achieved the desired result with satisfying performance.

Two techniques of image augmentation, image rotation and horizontal flipping on the training dataset are performed before feeding it to the network for training. After the training process the model was evaluated on several measures for statistical values. The predictions produced from the model on test images were ensembled using bagging and then post-processed using the thresholding technique to remove the greyish predictions around the predicted lesions.

6. REFERENCES

Badrinarayanan, V., Kendall, A.,.Cipolla, R. 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.*In IEEE Transactions on Pattern Analysis and Machine Intelligence,* 12(39), 2481-2495.

Brahmbhatt, P., S. N.Rajan, S. N. 2019. Skin Lesion Segmentation using SegNet with Binary Cross-Entropy. *International Conference on Artificial Intelligence and Speech Technology* (pp. 145-152). Delhi: Indira Gandhi Delhi Technical University for Women.

Cha, K. H.,Hadjiiski, L., Samala, R. K. et al. 2016. Urinary Bladder Segmentation. *Med Phys, 43*(4),1882–1896. Retrieved July 2019, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4808067/.

Codella, N., G. D. 2017. *Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)*. RetrievedJuly 2019, from https://challenge.kitware.com/#challenge/583f126bcad3a51cc66c8d9a.

Girshick, R. 2015. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV).* Santiago, Chile.

Goldbloom, A. 2010. *Kaggle*. (Alphabet Inc.) Retrieved October 2019, from https://www.kaggle.com/.

Gutman, D. 2016. *Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)*. Retrieved July 2019, from https://arxiv.org/abs/1605.01397.

Ioffe, S., Szegedy,C. 2015. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. Retrieved July 2019, from https://arxiv.org/abs/1502.03167.

Kendall, A., Badrinarayanan, V.,Cipolla, R. 2019. SegNet. University of Cambridge, 2015. Retrieved October 2019, from http://mi.eng.cam.ac.uk/projects/segnet/.

Lopez, A. R.-i.-N. 2017. Skin lesion classification from dermoscopic images using deep learning techniques. In *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on* IEEE.pp. 49-54.

MdZahangir Alom, M. H.,Yakopcic, C.,Taha, T. M.,Asari, V. K. 2018. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. (Cornell University) Retrieved July 2019, from https://arxiv.org/abs/1802.06955v5.

Melanoma: Statistics | Cancer.net. 2019. American Society of Clinical Oncology (ASCO). Retrieved July 2019, from American Society of Clinical Oncology (ASCO).

Mendonça, T.,Ferreira, P. M., Marques, J. S., Marcal, A. R., Rozeira,J. 2013. PH2 - A dermoscopic image database for research and benchmarking. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5437-5440.

Perez, F., Vasconcelos, C., Avila, S., Valle, E. 2018. *Data Augmentation for Skin Lesion Analysis*. (Cornell University) Retrieved July 2019, from https://arxiv.org/abs/1809.01442.

Ronneberger, O., Fischer, P., Brox, T. 2015. U-Net: Convolutional Networks for BiomedicalImage Segmentation.Retrieved July 2019, from https://arxiv.org/pdf/1505.04597.pdf.

WHO. 2019. *WHO | Skin Cancer*. (World Health Organisation) Retrieved July 2019, from https://www.who.int/uv/faq/skincancer/en/index1.html.

Yuan, Y., Chao, M. C., Lo, Y-C. 2017. Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance. *IEEE Transactions on Medical Imaging*. 36(9), 1876 – 1886.