UNIVERSITY OF THESSALY

DEPARTMENT OF COMPUTER SCIENCE AND BIOMEDICAL INFORMATICS

MEDICAL IMAGE SEGMENTATION WITH ARTIFICIAL NEURAL NETWORKS: A COMPARATIVE STUDY

A PROJECT FOR THE LESSON
MEDICAL DECISION SUPPORT SYSTEMS

MADE BY
[LAZAROS KONSTANTINOS-PANAGIOTIS, 1639]

PROFESSOR
[DIMITRIS IAKOVIDIS]

LAMIA, 2021-2022

Medical Image Segmentation Using Deep Neural Networks

Konstantinos Lazaros *DIB*, *UTH*

Abstract— Semantic image segmentation is the task of partitioning an image into many different segments. It is essentially the process of labeling each pixel of an image with a corresponding class. The goal of this process is to extract semantic information(high level information) from an image (i.e: detecting where objects of interest are located in an image). This paper is a comparative study between two neural network models (Unet and DoubleUnet) which are used to perform semantic segmentation on medical images.

I. INTRODUCTION

In recent years there are a lot of modern semantic segmentation architectures which have managed to surpass other more classic ones in terms of performance and efficiency. Some of the most recent segmentation architectures include, ResUnet++, HyperDenseNet, deepmedic (deepmedic has done incredibly well in many different medical image segmentation challenges) as well as doubleUnet.

Medical image segmentation is the task of assigning a label to each pixel in a medical image in order to highlight an object of interest within the image. It is an extremely important task which is essential for creating computer aided medical support systems for disease detection and/or therapy planning. It allows clinical doctors to focus on specific areas of interest which are associated with a disease and extract semantic information which will lead to a more precise diagnosis. Convolutional Neural Networks (CNNs) have shown exquisite performance in tasks of medical image semantic segmentation.

There have been many different CNN approaches/architectures which have been designed with medical image segmentation in mind. U-net is one of the first popular architectures that was proposed for medical image segmentation. It's architecture looks like the letter "U" hence it's name. It consists of 414.404 parameters and it is essentially a combination of an encoder and a decoder network.

The encoder is the first part of the model; it consists of convolution blocks as well as maxpool downsampling blocks in order to extract features (create feature maps) from the input image. The decoder is the second part of the model; it consists of upsampling and concatenation blocks which are followed by regular convolution layers. It projects the low resolution features that have been extracted by the encoder in higher resolutions in order to create a dense classification.

What made U-net differ from other CNNs that where designed for image segmentation at the time, was the fact that the feature maps which where used for contraction where later used to expand an image vector and create a segmented image. This concept ensures better preservation of structural integrity of the image, which means that distortion is greatly reduced. There have been different variations of this architecture over the years (i.e: VGG-19; it's 19 layers deep).

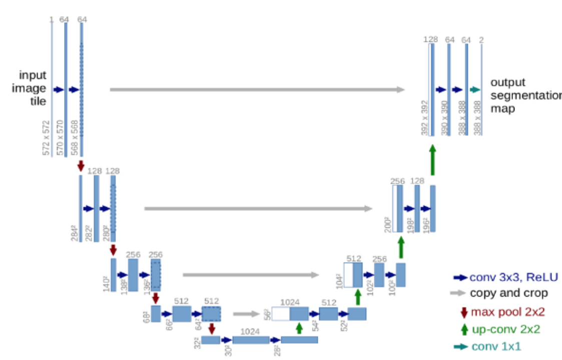

Fig. 1. Unet architecture

Another quite recent architecture which has been proposed for medical image segmentation is doubleUnet. DoubleUnet has been inspired by the success of U-net and it's early counterparts. It is for all intents and purposes one of U-net's more advanced successors. It consists of 29.303.426 parameters and is made up of two U-nets that are joined together. It's goal is to enhance segmentation performace by making several radical changes to the classic U-net architecture.

It starts with a VGG-19 encoder which is then followed by a decoder. The input image is fed into the first modified U-net which generates a prediction (segmented version of the image; a mask). Then the input image is multiplied with the prediction and the product is fed into the second modified U-net. The second U-net takes advantage of all the information that has been extracted by the first sub-network in order to create a second prediction. In the end, the two predicted masks are concatenated in order to get the final prediction.

This paper is a comparative study between two CNN architectures; The classic U-net and it's more up-to-date advanced derivative; The DoubleUnet. These two models have been evaluated on the CVC dataset. It medical image segmentation dataset with colonoscopy images. The masks included in the dataset depict the areas where polyps are located in the original images.

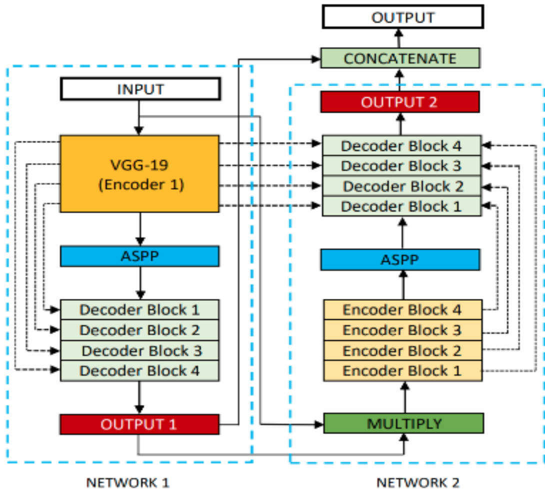

Fig. 2. DoubleUnet architecture

II. METHODOLOGY

- Before training and evaluating the two models, data processing takes place. More specifically the images from the given dataset are split into three different sets (same happens with their corresponding masks); a training dataset, a validation dataset and a test dataset. The training set is used to train the two models, while the validation set is used to give an estimate of each model's performance during training. The test set will be used in order to evaluate each model once training is complete.
- After splitting the images into three different categories, the training set images and their corresponding masks are augmented. Data augmentation is the process of increasing the amount of data available by adding slightly different copies of already existing data or by creating new, synthetic data. The training set images go through a number of different transformations in order to make this process work. Some of these transformations are geometric and some are essentially conversions from one colorpace to another (i.e: rgb to hsv). It should be noted that the colorspace conversions do not have an visible effect on the each image's corresponding mask whereas geometric transformations (i.e: flip, rotate, etc) do.
- Once data preprocessing is complete, building and training each network takes place. The last step is to evaluate each model and compare their results. It should be noted that due to limited computational sources, training doubleUnet was not possible due to it's complexity. Thus only Unet has been trained for 20 epochs and with a batch size of 8.
- Despite this, the code needed for building and training doubleUnet is also available in the project's github repository, in case anyone (with access to the required computational power) wants to build and train doubleUnet from scratch. Because of this a pretrained doubleUnet provided by it's original developers has been used for evaluation and comparisson against the "home-trained" unet model.

III. RESULTS

At first, the two models are evaluated on the test set that was created during the data preprocessing stage from the CVC dataset. The metrics used for evaluation are;

1) Accuracy: one of the most commonly used evaluation metrics. In this case it is the number of pixels that have been correctly classified. Accuracy is calculated by the following formula;

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

2) Precision: it is the ratio of correctly predicted positive observations to the total number of predicted positive observations. It's a measure of how pure positive detections are in relation to the ground truth. Precision is calculated by the following formula;

$$precision = \frac{TP}{TP + FP}$$

3) Recall: it is the ratio of correctly predicted positive observations to all observations in the actual class. It's a measure of how complete positive detections are in relation to the ground truth. Recall is calculated by the following formula;

$$recall = \frac{TP}{TN + FN}$$

4) IoU: it is also known as the Jaccard index. It is a metric which is used to quantify the overlap between the predicted mask and the ground truth. It is closely related to the dice coefficient which is commonly used as a loss function for training neural networks for image segmentation tasks. IoU is calculated by the following formula;

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

The evaluation results for each model are presented in the matrix below along with some result images for comparison.

Model metrics on CVC dataset evaluation				
Model	Accuracy	Precision	Recall	IoU
Unet	0.785	0.058	0.086	0.090
Wnet	0.994	0.993	0.927	0.966

The two trained models have also been evaluated on images taken from the KID dataset (provided by Dr. Dimitris Iakovidis et al.). The results from this evaluation process can be seen below.

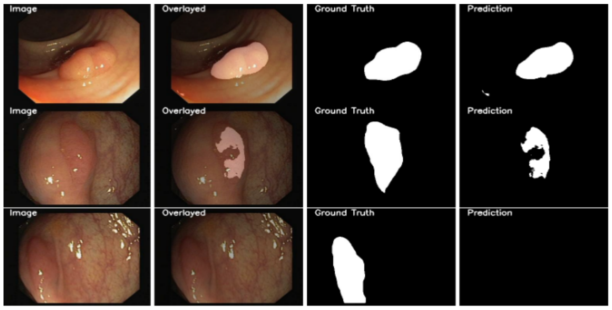

Fig. 3. Unet CVC-dataset evaluation sample images

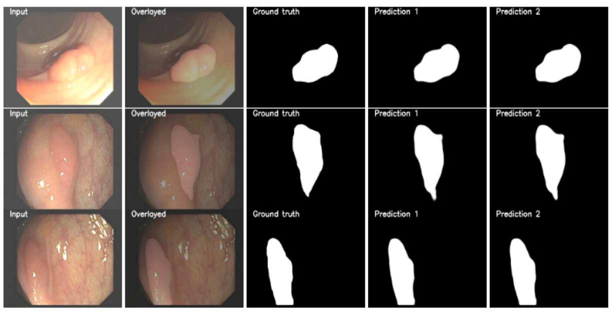

Fig. 4. DoubleUnet CVC-dataset evaluation sample images

Model metrics on KID dataset evaluation				
Model	Accuracy	Precision	Recall	IoU
Unet	0.170	0.998	0.361	0.102
Wnet	0.259	0.990	0.146	0.428

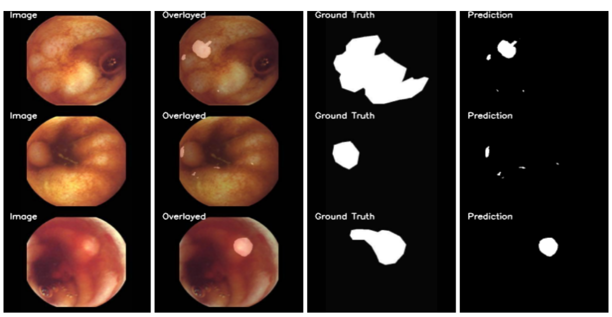

Fig. 5. DoubleUnet KID-dataset evaluation sample images

It is obvious that in both cases, doubleUnet manages to score better predictions than it's predecessor. While in KID's case both models don't make as accurate predictions (compared to CVC's predictions), their performance could possibly be improved by fine tuning their parameters on the new dataset through transfer learning, a task which will possibly be conducted for further testing in the near future.

IV. CONCLUSIONS

This is a comparative study between Unet and DoubleUnet, two neural network architectures that have been designed with the goal of image segmentation in mind. It is clear that doubleUnet has performed better in both evaluation cases than unet.

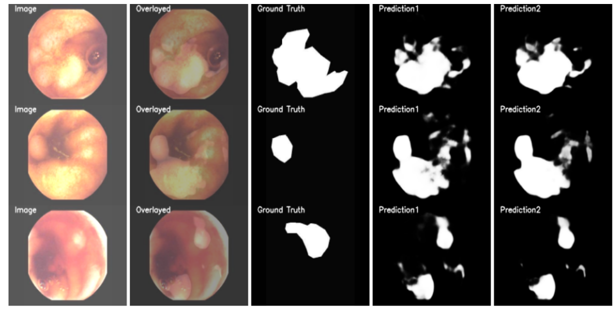

Fig. 6. DoubleUnet KID-dataset evaluation sample images

Aside from fine tuning the two models through transfer learning for making better predictions on the kid dataset, transfer learning could also be used in order to fine tune the two models for the task of semantic segmentation on ct scans/chest x-rays of covid-19 patients to determine the area of the lungs that has been affected by the disease. These goals will be part of a future study.

ACKNOWLEDGMENT

I would like to thank;

- Dr. Dimitris Iakovidis: Professor, Dept of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece..
- Dimitra-Christina Koutsiou: Ph.D. Student, Dept of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece.

for their constant encouragement towards the realization of this work.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Proceedings of International Conference on Medical image computing and computer-assisted intervention (MICCAI), 2015, pp. 234–241.

[2] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard et al., "Tensorflow: A system for large-scale machine learning," in Proceeding of USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2016, pp. 265–283.

[3] Fernández-Esparrach, G., Bernal, J., López-Cerón, M., Córdova, H., Sánchez-Montes, C., de Miguel, C. R., Sánchez, F. J. (2016). Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. Endoscopy, 48(09), 837-842.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2015, pp. 3431–3440. Volume 31– No.8, October 2011

[5] A. Koulaouzidis, D. K. Iakovidis, D. E. Yung, E. Rondonotti, U. Kopylov, J. N. Plevris, E. Toth, A. Eliakim, G. W. Johansson, W. Marlicz, G.W. Johansson, W. Marlicz, G. Mavrogenis, A. Nemeth, H. Thorlacius, G.E., Tontini, "KID Project: an internet-based digital video atlas of capsule endoscopy for research purposes," Endoscopy International Open, vol. 5, no. 06, pp. E477–E483, 2017.

[6] D.Jha, M. A. Riegler, D. Johansen , P. Halvorsen, H. D. Johansen, "DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation" IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), pp. 558-564, 2020.

[7] Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics, 43, 99-111.
[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2015, pp. 3431–3440.
[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in IEEE conference on computer vision and pattern recognition (CVPR), 2009, pp. 248–255.