Semantic Segmentation using U-Net architecture and its performance evaluation

1_{st} Muhammad Ather Hussain
dept. Electrical Engineering
Sukkur IBA University
Sukkur, Pakistan
matherhussain.mef18@-iba-suk.edu.pk

2^{nd} Dr. Gulsher Ali
dept. Electrical Engineering
Sukkur IBA University
Sukkur, Pakistan
gulsher@iba-suk.edu.pk

***Abstract*—In this paper, semantic segmentation has been performed using machine learning. The Fully Convolutional Networks and U-Net algorithms were studied, and finally U-Net architecture was chosen to perform semantic segmentation in python. The custom dataset was prepared that included four different geometrical shapes. Algorithms outperformed in terms intersection over union and F_1 score.**

***Keywords*—Machine learning, Semantic Segmentation, U-NET.**

I. Introduction

During this decade, deeper convolutional neural networks proved to best performed in various classification domains. Since this type of neural networks has been applied before but limitation of training and test dataset didnt allow researchers to further improve the network. The Krizhevsky et al [1] was able to train using supervised learning on the large ImageNet dataset with 1 million images. Since then, it made also possible to even other deeper and larger networks to be trained in same way.

The conventional use of convolutional networks is classifying the different tasks, where predicted output is label of single image. However, in different applications especially in biomedical image analysis where, it is necessary to specify and localize that desired output. It was only possible to using pixels level approach so researcher started to explore deep into image segmentation.

Since the biomedical images has are not publicly available due to patient's security reasons. So, Ciresan et al [2] used sliding-window setup to predict every pixel by defining region of interest (ROI) or patch around that pixel as input. First, this network was localizing at input image and second that training in terms of patches was greater than total training set. Thus, this network won the EM segmentation challenge at ISBI 2012 by a large margin.

Since Ciresan et al [2] has two disadvantages. Network is run for every individual patch, so it is quite slow, and there is dependency issue where patches overlaps. Another is that patches with larger size increase the numbers of max pool layer that correspondingly degrades the localization accuracy whereas, smaller one gives poor contextual information.

This paper focuses to study two different image segmentation algorithms knowns as Fully Convolutional Neural Network (FCN) and U-Net. Further, both architectures are compared to each other which suggests U-Net over Finally, U-Net is trained on custom dataset to evaluate its performance in terms of Intersection over Union (IU) and F_1 score. Novelty of this architecture is that it uses very few training samples and gives precise segmentation.

II. Background

A. Image Segmentation

Image segmentation is referred as the process of decomposing an image into multiple segments. Image segmentation is typically used to locate objects and boundaries in images.

1) Semantic Segmentation: Semantic Segmentation partitions the image in parts where each part is classified into one of the pre-determined classes. It assigns same label to those pixels which belongs to same class.

2) Semantic Segmentation: Instance segmentation classifies individual objects within a scene, regardless of if they are the same type. This method allows to computer to identify different objects of same class.

III. Methods

A. Fully Convolutional Networks (FCN)

FCN is considered as most widely used network for semantic segmentation. Usually convolutional netwroks such as Alexnet, Googlnet or VGG-16 [3] classify an image to a single label as shown in Fig 1. While, the FCN labels the class for each individual pixel of an input image and reconstructs output of same size as input can be seen in Fig 2. It decomposes the intermediate layers like max pool and convolutional layers to 1/32th of input and then, it makes prediction of pixel class in image. In reconstructing process, models uses upsampling and deconvolutional layers. Since there are different version of FCN as FCN-32, FCN-16 and

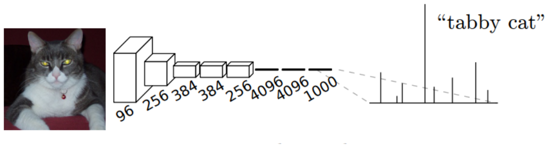

Fig. 1. Convolutional Network.

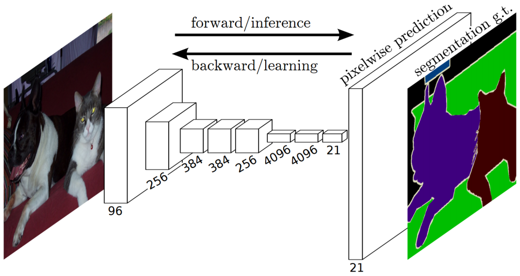

Fig. 2. Fully Convolutional Network.

FCN-8 but most of the time FCN-8 is preferred due to accuracy in terms of localization and contextual information.

B. U-Net

The U-Net architecture is modified version of Fully Convolutional Network (FCN). It is designed as U shape so, by it's structure it is called U Network or in short as U-Net. Since it uses data augmentation so it was specifically designed for the segmentation of biomedical images where a limited annotated data is usually available.
Since it is built upon FCN so it is superior over FCN becuase of its:

- Symmetric structure.
- Sum replaced by concatenation in skip connection between upsampling and downsampling paths.

C. U-Net Network Architecture

This network architecture illustrated in Fig. 3 is comprised of 23 layers where it includes the contracting and expanding path on left and right side respectively [4]. The contracting path is based on conventional convolutional network. It applies unpadded 3x3 convolutions two times where each convolution is followed by Rectified Linear Unit (ReLu). After that a 2x2 max pooling operation with 2 strides is performed for downsampling. For every downsampling operation, it doubles the feature channels. On the right side, the expanding path is comprised of an upsampling of features followed by a 2x2 convolution that which halves the no. of feature channels. After that concatenation with the corresponding cropped feature map from contracting path is evaluated and two 3x3 convolutions followed are computed each followed by ReLu. Due to the edge pixel loss in for each convolution, the cropping the image became necessary. For mapping each 64-component feature vector, a final layer of 1x1 convolution is used.

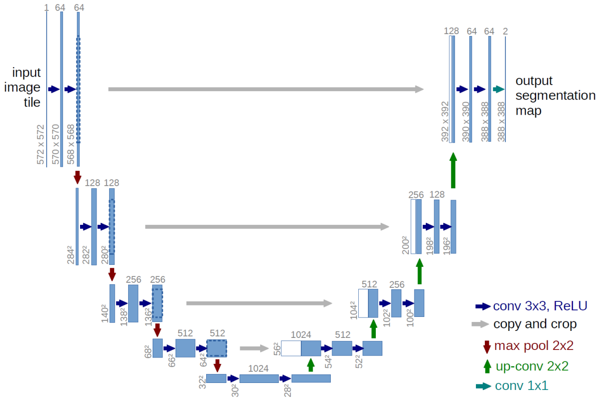

Fig. 3. U-Net architecture.

D. Dataset

The dataset used in this paper is manually created by Seth Adams and can found on Github [5]. He used the four different shapes: Star, Circle, Square and Triangle. Due to invariancy in dataset, data augmentation is applied on by capturing images at different angles and positions. The data was annotated using labelme python library and .json is created individually for every image in training dataset.

E. Implementation

The proposed model is implemented using Keras library built on TensorFlow 1.13 framework. Training was performed using Corei7 with processing speed of 3.60GHz and 8GB DDR3 RAM. The model runs total 500 epochs with 40 batches in each epoch where it took around 16 hours to finish the process. The model is solely trained for binary segmentation.

IV. Results and Discussion

The proposed architecture outperformed on the custom dataset. It was tested on different test images to perform semantic segmentation. Since performance was evaluated using IU and F1 score because accuracy metric didnt not always ensure the perfect health of an algorithm when there is more of ratio of background in image than the targeted class present. IU and F_1 score was calculated for test images using equ. 1 and equ. 2.

$$IU = \frac{\text{area of overlap}}{\text{area of union}} \tag{1}$$

and

$$F_1 Score = \frac{2 * \text{area of overlap}}{\text{total pixels combined}} \tag{2}$$

Since IU, F_1 score and predicted output as shown in Fig. 6 convince that model is robust and performed excellently on different input inputs.

TABLE I
PERFORMANCE METRICS

SN	*IU*	*F_1 Score*
1	0.94647795	0.9702966
2	0.98537576	0.98982245
3	0.77875095	0.798520571

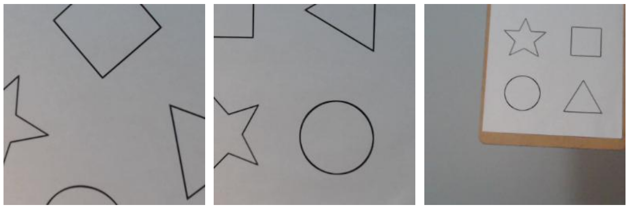

Fig. 4. input images

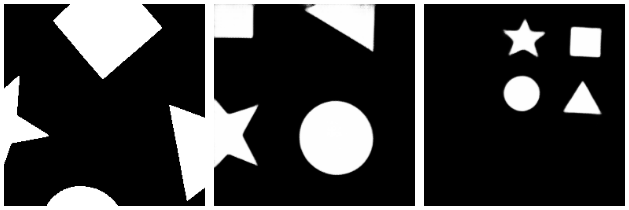

Fig. 5. Ground truth images

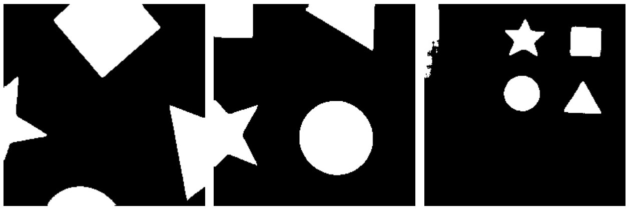

Fig. 6. Predicted semantic maps

V. CONCLUSION & FUTURE WORK

U-Net implemented in paper is widely adopted for semantic segmentation purpose. It is mainly preferred for medical image segmentations. This paper also ensures that it can also be used in custom datasets. Study suggests that the performance of this method can also be improved using different non-linear function such as Exponential Linear Unit (ELU), tanh and leaky ReLu etc. In future, this work can also be extended to perform multi-class segmentation.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[2] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[3] 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*, 2015, pp. 3431–3440.

[4] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[5] S. Adams, “Semantic shapes: Custom semantic segmentation tutorial/ pipeline,” https://github.com/seth814/Semantic-Shapes.git, 2013.