GT-Net: A Deep Learning Network for Gastric Tumor Diagnosis

Yuexiang Li†
Youtu Lab
Tencent
Shenzhen, China
vicyxli@tencent.com

Xinpeng Xie†
Computer Vision Institute
Shenzhen University
Shenzhen, China
xixipeng2017@email.szu.edu.cn

Shaoxiong Liu
Sixth People’s Hospital of Shenzhen
Shenzhen, China
liusx2008@sina.com

Xuechen Li
Computer Vision Institute
Shenzhen University
Shenzhen, China
timlee@szu.edu.cn

Linlin Shen*
Computer Vision Institute
Shenzhen University
Shenzhen, China
llshen@szu.edu.cn

Abstract—Gastric cancer is one of the most common cancers, which causes the second largest number of deaths in the world. Traditional diagnosis approach requires pathologists to manually annotate the gastric tumor in gastric slice for cancer identification, which is laborious and time-consuming. In this paper, we proposed a deep learning based framework, namely GT-Net, for automatic segmentation of gastric tumor. The proposed GT-Net adopts different architectures for shallow and deep layers for better feature extraction. We evaluate the proposed framework on publicly available BOT gastric slice dataset. The experimental results show that our GT-Net performs better than state-of-the-art networks like FCN-8s, U-net, and achieved a new state-of-the-art F1 score of 90.88% for gastric tumor segmentation.

Keywords—Gastric Tumor, Fully Convolutional Network, Segmentation.

I. INTRODUCTION

Gastric cancer (GC) is the most common cancer and it causes the second largest number of cancer related deaths globally. GC is characterized as an aggressive malignancy, which is very tough to be detected at an early stage. The pathological examination is one of the most commonly used diagnosis method for GC, which provides definitive disease diagnoses to guide patient treatment and management decisions. However, pathological inspection requires pathologists to manually annotate the gastric tumor in gastric slices, which is expensive, laborious and time-consuming. Over the past several decades there has been increasing interest in developing computational methods to assist the analysis of microscopic images in pathology.

Since the eight-layer AlexNet [1] won the ILSVRC competition in 2012, convolutional neural network (CNN) attracted increasing attentions from the community. Increasing number of researchers try to apply deep learning technique to pathological images analysis. For examples, Oikawa et al. [2] presented a hybrid GC detection system for pathological images, which combines the support vector machine (SVM) and CNN. Xu et al. [3] presented a deep CNN based framework for identifying epithelial and stromal regions in histopathological images. The image was separated into small patches, CNN were employed to identify the epithelial and stromal tissues. Wang et al. [4] presented a deep learning based breast cancer detection method. Similar to Xu’s work, the whole image was divided into patches. CNN was employed for patch classification. However, the CNN is mainly proposed to address the classification task, which is not suitable for segmentation.

In 2015, Long et al. [5] proposed the fully convolutional network (FCN) for semantic segmentation. The main idea under this model is to apply the contemporary classification networks (AlexNet [1], VGG net [6], and GoogLeNet [7]) to segmentation task. Due to the effective and efficient segmentation results produced by FCN, the model has been used for various medical applications. Ronneberger [8] developed a symmetric structural FCN, U-net, for microscopy image segmentation. Chen [9] proposed a FCN-based network for the segmentation of glands. Albarqouni developed a novel deep learning network, namely AggNet, for the mitosis detection in breast cancer histology images [10]. Nan et al. [11] uses active learning to incrementally train the U-net for the segmentation of gastric slices. However, most of the networks for pathological images were based on the VGG-like structure, i.e. straightforward without shortcut branches for down-sampling process. As the deep learning model developed, multi-branches networks, e.g. Inception [7, 12] and ResNet [13], have gradually surpassed straightforward networks and become the main-stream model for researchers.

†Equal contribution; *Corresponding author

The work was supported by Natural Science Foundation of China under grants no. 61672357, 61702339 and 61702337, the Science Foundation of Shenzhen under Grant No. JCYJ20160422144110140, and the China Postdoctoral Science Foundation under Grant No. 2017M622779.

2018 IEEE 30th International Conference on Tools with Artificial Intelligence
2375-0197/18/$31.00 ©2018 IEEE
DOI 10.1109/ICTAI.2018.00014
In this paper, we proposed a novel multi-branches fully convolutional network for the segmentation of gastric tumor in pathological slices. The contribution of this paper is listed as below:

1) We proposed a multi-branches fully convolutional network, namely GT-Net, adopting different architectures for shallow, deep and up-sampling layers, for gastric tumor segmentation.

2) Experimental results on publicly available gastric slice dataset demonstrate the outstanding segmentation performance of our GT-Net, which outperforms several well-known FCNs and achieves a new state-of-the-art segmentation accuracy.

II. RELATED WORKS

A. Atrous Convolution

Atrous convolution (or dilated convolution) was employed for deep learning based semantic segmentation approaches in many recent researches [14]. Compared to common convolution layer, the convolutional kernels in atrous convolution layer have ‘holes’ between their parameters, which enlarge the receptive filed and preserve the number of trainable parameters. The size of ‘holes’ inserted to the parameters is decided according to the dilation rate (γ). As shown in Fig. 1, a smaller dilation rate results in a more compact kernel, i.e. the common convolution is with dilation rate = 1, while a larger dilation rate produces an expanded kernel, which can capture more context information from the feature maps in previous layer.

![Fig. 1. Example of atrous convolutions with different dilation rates. The purple squares represent the positions of kernel parameters.](image-url)

III. METHOD

A. Dataset

The recently published BOT gastric slice dataset is utilized in this work. The gastric dataset contains 560 gastric cancer slices and 140 normal slices. The slices were stained by hematoxylin-eosin (H&E) staining and captured using a microscope with a magnification factor of 20x. The resolution of gastric slice is 2048 x 2048. The tumor areas are partly annotated by the data provider. In our experiments, while 80% of gastric slices (normal and cancer) are randomly selected for network training, the remaining 20% slices are used for testing.

B. Pre-processing & Data augmentation

The resolution of original gastric images is too large for deep learning network to process directly. The gastric images in training and validation set were cropped to 224 x 224 patches for networking training and testing. For the normal gastric slice, we slide the window over the whole image to generate patches, while the window is only located over the cancer area for gastric cancer slice. As the BOT dataset is partly annotated, we asked the experienced pathologists to manually annotate the extracted patches, which make the dataset become fully-annotated for training and testing. To augment the training set, the training patches are rotated by 90°, 180° and 270°.

C. The proposed GT-Net

The flowchart of our GT-Net is presented in Fig. 2. The blue and red rectangles represent the convolutional layer and max pooling layer, respectively. To extract and fuse features, the proposed GT-Net adopts multi-scale module (MSM) for shallow layers and feature pyramid for deep layers. The extracted feature maps are then up-sampled back to the size of original image and yield the segmentation result using the up-sampling convolutional module (UCM). The Batch Normalization and ReLU layers are placed in front of each convolutional layer. The network optimization is supervised by binary cross-entropy loss and dice-coefficient loss.

![Fig. 2. Flowchart of GT-Net. The blue and red rectangles represent the convolutional and max-pooling layers, respectively. MSM and UCM stand for multi-scale module and up-sampling convolutional module, respectively.](image-url)

1) Multi-scale module (MSM): The feature maps for shallow layers have higher resolution than that for the deep layers, which may contain objects of different scales. As convolutional layer with single kernel size is unable to simultaneously detect multi-scale targets, we proposed the multi-scale module (MSM) for shallow layers to address the problem. Fig. 3 shows the architecture of multi-scale module.

![Fig. 3. The architecture of multi-scale module (MSM). The first MSM was taken as example.](image-url)
are concatenated for the next layer. To avoid the problem of feature explosion, a 3 x 3 convolutional layer is added to the end of MSM to fuse and compact the generated features.

2) Feature pyramid: The size of feature maps passed through three max-pooling layers becomes 28 x 28. To effectively use the information contained in the feature maps, a feature pyramid module is proposed. Fig. 4 shows the architecture of proposed feature pyramid module. The orange and green rectangles represent the average pooling (AP) and reshape layers, respectively. The proposed feature pyramid module fuses features under five different pyramid scales. The third parameter of 1 x 1 convolutional layers represents the number of convolutional filters. The outputs of convolutional layers are then transformed to the same shape, i.e. 28 x 28 x 26, by reshape layers. The number of channels for each average pooling branch is decided by averaging the number of channels of input feature maps (128/5≈26). The output of feature pyramid is produced by concatenating the features generated by different average pooling branches and the input feature maps.

The architecture of our feature pyramid module is similar to that of pyramid pooling module (PPM) proposed in [15]. However, we replace the bilinear interpolation operation in up-sampling step to dense up-sampling convolutional module (DUC) [16], which is able to capture and decode more detailed information.

3) Up-sampling convolutional module (UCM): To generate the final segmentation result, the 28 x 28 feature maps need to be up-sampled back to the size of original image. The dense up-sampling convolutional module (DUC) [16] is adopted for up-sampling in our GT-Net. To better classify the two categories, i.e. tumor and normal areas, the features maps corresponding to different categories should be separately processed. Hence, a depthwise separable convolutional layer [13] with depth = 2 is added to the typical DUC, which forms our up-sampling convolutional module (UCM). Table I listed the pipeline of the proposed UCM.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Type</th>
<th>Kernel size, number &amp; depth, dilation</th>
<th>Output size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>3x3, 256, 1, 1</td>
<td>28x28x256</td>
</tr>
<tr>
<td>2</td>
<td>SC</td>
<td>3x3, 512, 2, 2</td>
<td>28x28x512</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>3x3, 64, 1, 1</td>
<td>28x28x64</td>
</tr>
<tr>
<td>4</td>
<td>R</td>
<td>224x224x1</td>
<td>224x224x1</td>
</tr>
</tbody>
</table>

D. Implementation

The proposed framework is developed with Keras toolbox. The GT-Net is trained using four GPUs (GeForce GTX TITAN X, 12GB RAM) with a batch size of 64. The initial learning rate is set to 0.0001. ‘Adam’ [17] is used as the solver for Stochastic Gradient Descent. The network converges after 30 epochs of training.

E. Differences with existing frameworks

We noticed that the proposed MSM, Feature pyramid and UCM have similar architectures compared to the Inception blocks [7]/atrous spatial pyramid pooling (ASPP) [12], pyramid pooling module (PPM) [15] and dense up-sampling convolutional module (DUC) [16], respectively. We compare the proposed modules with the existing ones and list the differences as below:

Multi-scale module (MSM): The Inception blocks in Google Inception networks adopt common convolutions of different kernel sizes, i.e. 1 x 1, 3 x 3, 5 x 5, to achieve multi-scale feature extraction, while the proposed multi-scale module (MSM) employed atrous convolutions with different dilation rates, which are more computational efficient. The atrous spatial pyramid pooling (ASPP) concatenates atrous convolutions with dilation rates of 6, 12, 18 and 24. The dilation rates are so large that the receptive field of deeper layer exceeds the patch size and lots of padding zeros are involved for convolution calculation. Compared to ASPP, our MSM has deeper branches with smaller dilation rates, i.e. each branch has two-layers with decreased dilation rates of 1, 2 and 3, which can extract high level features with appropriate receptive field sizes.
• **Feature pyramid:** The proposed feature pyramid replaces the bilinear interpolation modules in original pyramid pooling module (PPM) with the dense up-sampling convolutional modules, which improves the capacity of capturing detailed information of our GT-Net.

• **Up-sampling convolutional module (UCM):** As the features maps corresponding to different categories, i.e., tumor and normal areas, should be separately processed, a depthwise separable convolutional layer with depth = 2 is added to the typical dense up-sampling convolutional module (DUC) to form our up-sampling convolutional module (UCM).

The proposed GT-Net is a patch-based deep learning framework, which can excellently deal with the incomplete annotation. It can be observed from Fig. 5 (c) and (f) that the GT-Net successfully detects the gastric tumor areas in the slices, which are not annotated in the ground truths. The pathologists from the hospital collaborating with us on this project have verified the results.

As the ground truths of BOT dataset are partly annotated and the pathologists only annotated the extracted patches, we adopt the patch-based F1 score to assess the framework performances in the following sections.

### IV. RESULTS

#### A. Evaluation criterion

F1 score is adopted as the metric for the evaluation of segmentation performance. Using the provided ground truths, the F1 score can be calculated via:

\[
F1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]  

(1)

\[
\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}
\]  

(2)

where the values of TP (number of true positives), FP (number of false positives) and FN (number of false negatives) were computed according to the definitions proposed in previous work [18].

#### B. Segmentation results

The publicly available BOT gastric slice dataset is partly annotated, which means that only part of gastric tumors have been marked in the ground truths, as shown in Fig. 5 (b) and (e).

![Fig. 5. Segmentation results on BOT gastric slices.](image)

The performance evaluation between proposed modules and the existing ones will be presented in the next section.

### C. Performance analysis

We evaluate the improvements generated by the proposed feature pyramid and up-sampling convolutional module (UCM) on BOT testing set. The results are presented in Table II and Table III.

#### TABLE II. F1 scores of GT-Net with/without feature pyramid (FPy) (%)

<table>
<thead>
<tr>
<th></th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>without FPy</td>
<td>89.15</td>
</tr>
<tr>
<td>with PPM</td>
<td>90.15</td>
</tr>
<tr>
<td>with FPy</td>
<td>90.88</td>
</tr>
</tbody>
</table>

The feature pyramid module (FPy) enables the GT-Net to detect multi-scale features from feature maps, which can boost the segmentation accuracy. As shown in Table II, the segmentation performance of the framework significantly increases from 89.15% to 90.88% with the feature pyramid. Due to the use of dense up-sampling module, the proposed feature pyramid involves more detailed information for the recognition of gastric tumors. As a result, the proposed feature pyramid outperforms the pyramid pooling module (PPM) with an improvement of 0.73%.

#### TABLE III. F1 scores of UCM using common/separable convolutional layer (%)

<table>
<thead>
<tr>
<th></th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCM with C (DUC)</td>
<td>90.06</td>
</tr>
<tr>
<td>UCM with SC</td>
<td>90.88</td>
</tr>
</tbody>
</table>

For the up-sampling convolutional module (UCM), we mainly assess the improvement generated by the separable convolutional layer (SC). Table III compares the performances of UCM using common convolutional layer (C) and separable convolutional layer (SC) (Noted that the UCM with common convolution can be seen as the original DUC). The separable convolutional layer enables the UCM to separately process the feature maps belonged to different categories, i.e. gastric cancer area/background, which is useful for the detection of gastric areas. Hence, it can be observed from Table III that the UCM with separable convolutional layer achieves a better F1 score of 90.88%, which is 0.82% higher than that of UCM using common convolutional layer (C).
D. Comparison with benchmarking algorithms

To evaluate the segmentation performance of our GT-Net, several famous fully convolutional networks, i.e., FCN-8s [5], SegNet [19], U-net [8] and fully convolutional ResNet-50/101 (FCRN-50/101) [13], were involved for comparison. The F1 scores of different algorithms on BOT testing set are listed in Table IV.

It can be observed from Table IV that our GT-Net achieves an F1 score of 90.88%, which outperforms all of the listed well-known deep learning frameworks. The runner-up is the U-net (89.07%), which is 1.81% lower than that of GT-Net. Due to the overfitting problem, the F1 scores of fully convolutional networks transformed from ultra-deep ResNet (FCRN-50/101) are relatively low among the listed benchmarking algorithms, which are 85.75% and 86.11%, respectively.

**TABLE IV. F1 SCORES OF DIFFERENT DEEP LEARNING FRAMEWORKS ON BOT TESTING SET (%)**

<table>
<thead>
<tr>
<th>Framework</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN-8s [5]</td>
<td>86.76</td>
</tr>
<tr>
<td>SegNet [19]</td>
<td>87.67</td>
</tr>
<tr>
<td>U-net [8]</td>
<td>89.07</td>
</tr>
<tr>
<td>FCRN-50 [13]</td>
<td>85.75</td>
</tr>
<tr>
<td>FCRN-101 [13]</td>
<td>86.11</td>
</tr>
<tr>
<td>GT-Net (ours)</td>
<td>90.88</td>
</tr>
</tbody>
</table>

Nan et al. proposed an active-learning based training scheme (ALT) for U-net and achieved the state-of-the-art performance for gastric tumor segmentation [11]. We adopt the ALT training scheme for our GT-Net to further boost its performance. Table V listed the F1 scores of ALT-trained U-net and the proposed GT-Net. As shown in Table V, the F1 score of GT-Net increases from 90.88% to 91.01% using ALT training scheme, which is 1.17% higher than that of U-net + ALT.

**TABLE V. F1 SCORES OF DIFFERENT ALT-TRAINED NETWORKS (%)**

<table>
<thead>
<tr>
<th>Framework</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-net + ALT [11]</td>
<td>89.84</td>
</tr>
<tr>
<td>GT-Net (ours)</td>
<td>90.88</td>
</tr>
<tr>
<td>GT-Net + ALT (ours)</td>
<td>91.01</td>
</tr>
</tbody>
</table>

It is worthwhile to mention that the performance of our GT-Net without ALT, i.e. 90.88%, still outperforms that of U-net + ALT, i.e. 89.84%, which demonstrates the excellent segmentation capacity of proposed network architecture.

V. CONCLUSION

In this paper, we presented a deep learning framework, i.e., GT-Net, for gastric tumor segmentation. The novel modules, i.e. Multi-scale module (MSM), Feature pyramid and Up-sampling convolutional module (UCM), are proposed for GT-Net for feature extraction, feature fusion and up-sampling, respectively. The GT-Net has been evaluated on publicly available BOT gastric slice dataset. The experimental results show that the GT-Net outperforms several well-known deep learning frameworks in the task of gastric tumor segmentation, and a new state-of-the-art F1 score of 90.88% was achieved.

**REFERENCES**


