
Precise Augmentation and Counting of Helicobacter Pylori in Histology Image

Yufei Cui

School of Computer Science
McGill University
yufei.cui@mail.mcgill.ca

Yixin Chen

Bingli Research
yxchen@ibingli.com

Zhifeng Shuai*

Zhejiang Hospital
zhifengshuai@163.com

Fang Peng*

Zhejiang Hospital
zjyblkzy@163.com

Yanbo Lv

Zhejiang Hospital
lvyanbo2021@163.com

Luoning Zheng

Zhejiang Hospital
lauren20210201@163.com

Xue Liu

School of Computer Science
McGill University
xueliu@cs.mcgill.ca

Antoni B. Chan

Department of Computer Science
City University of Hong Kong
abchan@cityu.edu.hk

Tei-Wei Kuo

School of Computer Science
National Taiwan University
ktw@csie.ntu.edu.tw

Chun Jason Xue

Department of Computer Science
City University of Hong Kong
jasonxue@cityu.edu.hk

Abstract

We study the precise counting of Helicobacter Pylori (HP), which is important for diagnosis of gastric cancer. The crowd counting technique is adapted for a precise quantitative analysis. The challenge of training an HP counting model lies in scarcity of labels. We use a DCGAN for the generative modelling of HP morphology and perform high-fidelity data augmentation. The comparative results show our method outperforms the object detection and semantic segmentation baselines. The proposed framework is potential useful in quantitative analysis of other bacteria in histology images. The dataset is available at <https://cyxhello.github.io/HPCDataset/>.

1 Introduction

The infection of Helicobacter Pylori (HP) is one of the most common bacterial infection and has been playing an important role for the diagnosis of gastric cancer [1]. The high density distribution of HPs over digital histology images and the complex morphology of single HP make annotation of segmentation mask time-consuming thus costly. Previous works use imprecise marks for annotation and weakly supervised learning for detecting the HP, which makes quantitative analysis (e.g., how many HPs are there in an image patch) impossible [2, 3]. This work study how to precisely count the number of HPs using machine learning, with a low amount of annotated images with only single dot as the annotation. We argue that the crowd counting pipeline is more suitable for the quantitative analysis of HPs, as it conducts a deep regression directly on the target number in an image [4].

*Corresponding authors

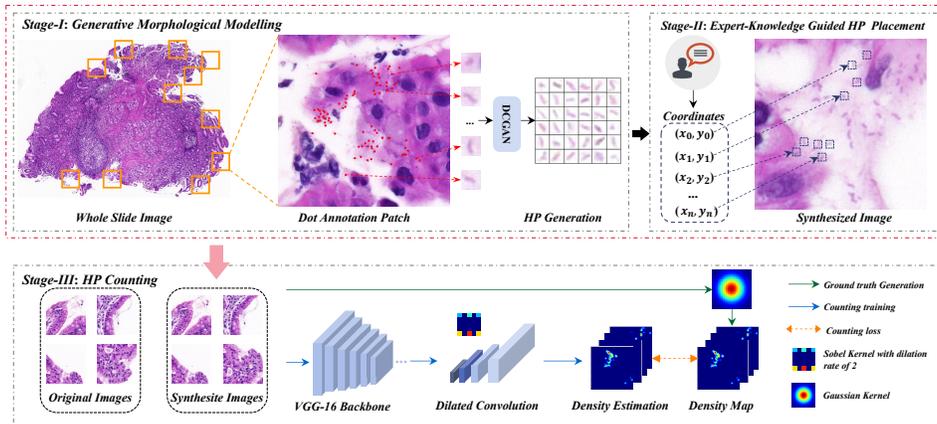


Figure 1: The framework of precise augmentation and counting of HP (PAC-HP).

Table 1: Statistics of image patches (1024x1024) of different datasets in this paper.

Training data	Number of images	Count		
		Min	Avg	Max
Real data	659	1	37	532
Synthetic data (abnormal)	1517	4	78	632
Synthetic data (normal)	3400	5	44	295
Real+Synthetic	2176	1	65	632

To solve the low data problem, we design a data generation pipeline to enrich the dataset. For generating high fidelity image, we use a generative model to capture complex the morphology of the HPs. The generated HPs are placed in original HP-infected or benign images based on rules from doctors. We merge the generated HPs pieces and the original images with Poisson blending. We improve the Poisson blending with a mixed gradient such that the HPs are better integrated with the background. For obtaining the precise numbers of HPs, we adapt the crowd counting techniques to the task of counting bacteria, dubbed as *HP counting*. The *HP counting* significantly outperforms the semantic segmentation which previous methods are based on. Moreover, the generative pipeline is shown to reduce the variance of *HP counting* by a large margin, which was a severe problem in the original crowd counting problem. The proposed framework (Figure 1) is potentially extensible to the quantitative analysis of other bacteria and useful in solving the general large variance problem in crowd counting.

2 Data and Methodology

We collect 57 H&E stained whole slide images (WSI) from 57 patients. The image patches with high-density HPs distributions are partitioned. HPs are annotated with single dots for saving the annotation cost. This process produces 659 image patches of resolution 1024x1024, with averaging HPs count 37. The detailed statistics are listed in Table 1. The dataset are available at HPCDataset.

Design philosophy. We observe that there are no much image patches (659) annotated as the annotation cost for one image is fixed and high. These images would be treated as the training data of the discriminative machine learning model, while the low data amount could raise issues like overfitting, low-robustness, et cetera. However, the number of annotated HPs is large 24,305, which we could leverage for the modelling of HP morphology by fitting a generative model. The main feature of PAC-HP is to control the generation and counting of *single HP*. As a result, accurate number and coordinates of generated HPs could be provided to the discriminative model for high performance. The HP placement step could generate a large amount data by manipulating different placement strategies and selecting different image patches as the background. By comparison, previous works only exploit the discriminative model with bounding boxes on the mass of HPs, which causes inaccurate estimation and makes quantitative analysis impossible.

Table 2: Performances of models trained on different data distributions.

	MAE			MSE		
	U-Net	Yolov5	PAC-HP	U-Net	Yolov5	PAC-HP
Real	19.522	9.119	11.233	26.215	12.957	15.063
Synthetic(HP-positive)	16.954	8.985	7.008	22.839	12.083	9.484
Synthetic(HP-negative)	24.767	11.606	13.867	32.821	15.854	16.902
Real+Synthetic	15.154	8.711	5.881	21.657	11.557	7.492

Generative morphological modelling. For precisely controlling the number of HPs generated, we propose to let the generative model fitted on image pieces. We crop each HP with a tiny square box 16×16 . These small pieces are re-scaled as the input of a Deep Convolutional Generative Adversarial Network (DCGAN)[5]. We use DCGAN to capture the complex morphology of HP and generate the new HP pieces by sampling.

Expert knowledge guided HP placement. We place the generated HPs in the image patches for synthesizing new image patches. This encapsulates two steps:

- Determining the coordinates of placed HPs. Doctors knowledge could be leveraged in this step. In this paper, we use a simplified rule based on density of HPs. We predict a series of plausible locations guided by pathology experts, which HP are usually locate at the mucosa and epithelium of gastric tissues. Since HP is usually densely distributed, the image is divided into several blocks (64×64). If there is only one HP in the block, then we randomly select 4/5 coordinates to place the HP.
- Blending HPs with the image background. We use Poisson Blending [6] for generating high-fidelity images. However, the blended images tend to meet edge disappearance problems when segmented HP fragments in the patch. Hence, we refine the mixing gradients. When the target gradient field f is stronger than the source gradient field g , we fuse the g and the f at a ratio of 3:7, and otherwise, the guidance field v is a gradient field taken directly from a source image. It preserves the morphology of HP more completely, see Fig. 2

HP counting. We adapt the crowd counting techniques [4] for a precise prediction of the number of HPs. First, the dot annotated ground truth images are replaced with a Gaussian blurred density map [7]. Second, a VGG16 network [8] is used to learn the mapping between the input image patch and the ground truth density map. The final count of HPs for one image is obtained by summation over the output density map from the neural network.

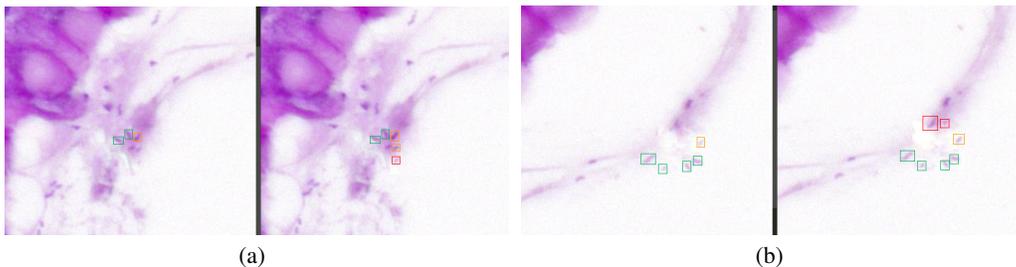


Figure 2: Illustration of Image blended by mixing gradient, refined mixing gradient. The green, orange, and red bounding boxes correspond to whole HPs, partial HPs, and missing HPs, respectively.

3 Experiments

For generative morphological modeling, we train DCGAN for 300 epoches with learning rate at $2e-4$ and batch size of 8. After careful quality assessment, we collect 6,000 single HP as candidate images. For HP counting, the model is pre-trained on ShanghaiTech B dataset [9] for 195 epochs and fine-tuned for another 65 epochs on our augmentation images. The model is trained using Stochastic gradient descent(SGD) with learning rate at $1e-6$ and a batch size of 2. Counting performance

of models in different settings is evaluated using MAE and MSE. Results of the experiments with synthetic data augmentation are shown in Table 2. The model trained with real data achieves a MAE of 11.233 and a MSE of 15.063. The model trained on synthetic data is capable of producing meaningful predictions when tested on real samples, demonstrating comparatively high performance. Combining the real samples with synthetic ones, we observe a prominent improvement in performance, which achieves the lowest MAE and MSE.

When compare against semantic segmentation(e.g. UNet [10]), which achieve a MAE of 19.522 and a MSE of 26.215. Our method outperform both in visualization quality and the ability to localize the HP. Yolov5-based detection model performs well on original real image, however, the benefits of image agumentation is limited, thus outperformed by the full PAC-HP.

4 Discussion

We propose the task of counting HP from whole slide images (WSIs). It necessitates thorough detection of morphology and distribution of HP. In this paper, we propose a complete pipeline for the HP counting. The stages include 1) Generative morphological modelling, 2) Expert-knowledge guided image placement based on Poisson editing and 3) HP counting. In addition, we present a large-scale HP counting corpus. To the best of our knowledge, it is the first such corpus. Extensive experiments conducted on real data demonstrate that our proposal outperforms the other detection approaches.

References

- [1] Sharon Zhou, Henrik Marklund, Ondrej Blaha, Manisha Desai, Brock Martin, David Bingham, Gerald J Berry, Ellen Gomulia, Andrew Y Ng, and Jeanne Shen. Deep learning assistance for the histopathologic diagnosis of helicobacter pylori. *Intelligence-Based Medicine*, 1:100004, 2020.
- [2] Kenta Ishihara, Takahiro Ogawa, and Miki Haseyama. Helicobacter pylori infection detection from gastric x-ray images based on feature fusion and decision fusion. *Computers in biology and medicine*, 84:69–78, 2017.
- [3] Yongquan Yang, Yiming Yang, Yong Yuan, Jiayi Zheng, and Zheng Zhongxi. Detecting helicobacter pylori in whole slide images via weakly supervised multi-task learning. *Multimedia Tools and Applications*, 79(35):26787–26815, 2020.
- [4] Yuhong Li, Xiaofan Zhang, and Deming Chen. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1091–1100, 2018.
- [5] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [6] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, pages 313–318. 2003.
- [7] Victor Lempitsky and Andrew Zisserman. Learning to count objects in images. *Advances in neural information processing systems*, 23, 2010.
- [8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [9] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 589–597, 2016.
- [10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.