Segmentation of Ascites on Abdominal CT Scans for the Assessment of Ovarian Cancer

Benjamin Hou
Radiology and Imaging Sciences, Clinical Center, National Institutes of Health
benjamin.hou@nih.gov

Manas K. Nag
Radiology and Imaging Sciences, Clinical Center, National Institutes of Health
manas.nag@nih.gov

Jung-Min Lee
Women's Malignancies Branch, National Cancer Institute, National Institutes of Health
leej6@mail.nih.gov

Christopher Koh
Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases
christopher.koh@nih.gov

Ronald M. Summers
Radiology and Imaging Sciences, Clinical Center, National Institutes of Health
rsummers@cc.nih.gov

Abstract

Quantification of the volume of ascites can be an accurate predictor of clinical outcomes to certain pathological setting, e.g., cases of ovarian cancer. Due to the properties of ascites being a liquid, accurate segmentation can be quite a challenging task. In this paper, we show that by tuning nnU-Net, a model that learns the heuristics of the data, it is possible to achieve state-of-the-art segmentation performance. Our trained model, was able to achieve a segmentation Dice score of 0.66, with 0.67 precision and 0.68 recall on pathological test cases. This is a distinct improvement over current state-of-the-art.

1 Introduction

Background: Ascites, by definition, is the accumulation of fluid in the peritoneal cavity that causes abdominal swelling. It is a condition commonly found in people who have cirrhosis of the liver (80% of cases), but can also be caused by other conditions, such as; cancer (10%), heart failure (3%), tuberculosis (2%), pancreatic disease (1%), or others (2%) [1]. The peritoneum, a sheet of tissue that lines the abdominopelvic cavity, surrounds the abdominal organs, i.e., stomach, bowels, liver, kidneys, and ovaries. It consists of parietal and visceral layers; the parietal peritoneum lines the abdominal and pelvic walls, and the visceral peritoneum wraps around the organs. The peritoneal cavity (i.e. the space in between) is where ascites manifests. Quantification of the volume of ascites can be an accurate indicator of disease severity, e.g., in the case of ovarian cancer, the volume of ascites at initial diagnosis of ovarian cancer is correlated with worse progression-free survival (PFS) and overall survival (OS) [9].

Segmentation of ascites can be a challenging feat, and is hindered by its properties of being a fluid. Unlike body organs, ascites cannot be inferred by a shape prior as it can be randomly distributed in the lower abdomen and pelvic region. Furthermore, the intensity profile of ascites is similar to other liquids in the abdominal region, e.g., water, urine, bile, etc. As CNNs overall are biased towards textures [2], it can potentially induce unanticipated false positives. Thus, it is preferable for 3D volumetric methods over 2D linear analysis for 2 purposes; to ensure connectivity in the axial plane and to incorporate anatomical location constraints.

Related Works: Segmentation and/or detection of ascites does not seem to be commonly explored in literature. Winkel et al [10] in 2019 proposed a detection based model to discriminate CT images with

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

and without fluid. Their model was able to achieve 85% sensitivity and 95% specificity in detection performance, however, additional manual analysis was required to quantify the volume in order to judge pathological severity. Ko et al. [5] in 2022 developed a method for automatic detection and quantification of ascites using 2D U-Net models (U-Net, Bi-directional U-Net and Recurrent Residual U-Net and Residual U-Net). In their approach, they found the Residual U-Net was best performing, achieving 96% sensitivity, 96% specificity, and 96% accuracy in segmentation performance. Nag et al. [6] in 2022 also developed a method for automatic detection and quantification of ascites, however, unlike Ko et al, their method was implemented in 3D to ensure segmentation connectivity in adjacent CT slices. The 3D model was provided anatomical location (in the form of a normalised score that spans from the chest to the lower extremities) of each voxel, achieving a Dice score of 0.65 after post processing.

With the advent of deep learning, and the introduction of U-Net [8], there has since been an upsurge of U-Net like variants for medical image segmentation. Unlike most variants in recent literature, nnU-Net [4] (No New U-Net) took a different approach of learning the data heuristics to determine the most appropriate training hyper-parameters instead of modifying the network architecture.

Contribution: Our model is able to distinguish ascites accurately from urine, water, bile, etc., achieving an average test dice of 0.66. Despite being trained only on contrast CT scans, the model is still able to segment ascites in non-contrast CT scans with high accuracy without the need of post-processing.

2 Methods

Our model is trained from a dataset consisting of both public (TCGA-OV [3]) and private (scans of patients from two clinical trials; NCT02203513 and NCT02484404) data. For both sources, labels of ascites were manually created by a trained expert under the supervision of a senior radiologist. As code and/or data is not public for related works, the methods in Ko et al. [5] were reimplemented in Tensorflow 2.0 + Keras and retrained on our dataset. A five-fold cross-validation was performed to assess the models' generalization ability. We further analyze the robustness of the models using the Dice metric, Jaccard Index, Precision, Recall, Sensitivity and Specificity.

3 Experiments and Results

Experiments were conducted on a mix of public and private in-house data. For the training dataset; 70 scans in total were taken from The Cancer Genome Atlas Ovarian Cancer Collection (TCGA-OV) [3], with an additional 30 scans taken from the NIH National Cancer Institute and NIH Clinical Center liver cirrhosis dataset. For the testing dataset; 10 scans were used from TCGA-OV dataset, 15 scans from the in-house liver cirrhosis dataset, and an additional 10 from an in-house ovarian cancer dataset. The training dataset consists of only intravenous contrast-enhanced CT scans, whilst the testing set is a mixture of contrast and non-contrast CT scans.

We compare our method to the most similar works in recent literature. For the 2D method in Ko et al. [5], their best performing network "2D Deep Residual U-Net (four residual blocks)" is chosen as a comparison point. 2D CT slices of size 512×512 were used to train the network without re-sampling to isotropic 1mm$\times$1mm. The network was trained for 200 epochs with a batch size of 8 and learning rate of 10^{-4} using the Adam optimiser. Training took approximately 2 days to reach convergence. For 3D Residual U-Net and BLE-U-Net [6], the networks were also trained using Adam optimiser with an initial learning rate of 10^{-4} and a batch size of 2 for 400 epochs. Random sub-patch sampling (patch_size = 64$\times$64$\times$64) was used for training, with sliding window inference (patch_size = 64$\times$64$\times$64 and stride = 60 allowing for 4px overlap) was used for testing. Training took approximately 12 hours to reach convergence. Both models were reimplemented and trained according to the descriptions provided by the original authors to the best of our abilities.

Our proposed method uses an nnU-Net [4] at its core. The network was trained using an SGD optimiser with an initial learning rate of 10^{-2} and a batch size of 2. Training took approximately 2 days to reach convergence. For all experiments, the CT intensity window was set at $[-175, 275]$ HU and subsequently intensity normalised to $[0, 1]$. The loss function is a combination of binary cross-entropy and soft Dice loss with equal weighting (Equation 1). All experiments were conducted on an NVIDIA DGX-1 machine with A100 GPUs. Table 1 shows the results of all methods.

$$\mathcal{L}(y, \hat{y}) = 1 - \frac{2 \cdot \sum_{i=0}^{N} y_i \cdot \hat{y}_i}{\sum_{i=0}^{N} (y_i + \hat{y}_i) + \epsilon} - \frac{1}{N} \sum_{i=0}^{N} \left(y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i) \right) \quad (1)$$

	F1 / Dice	Jaccard	Precision	Rec./Sen.	Specificity
2D Residual U-Net [5]	0.641 [0.621,0.661]	0.499 [0.476,0.522]	0.684 [0.657,0.711]	0.636 [0.610,0.662]	0.989 [0.988,0.990]
3D Residual U-Net	0.561 [0.547,0.575]	0.416 [0.402,0.430]	**0.704** [0.678,0.730]	0.509 [0.491,0.527]	**0.990** [0.989,0.991]
3D BLE-U-Net [6]	0.471 [0.433,0.509]	0.337 [0.304,0.370]	0.582 [0.510,0.654]	0.464 [0.386,0.542]	0.984 [0.981,0.987]
nnU-Net (proposed)	**0.658** [0.644,0.672]	**0.519** [0.503,0.535]	0.670 [0.646,0.694]	**0.683** [0.664,0.702]	0.988 [0.987,0.989]

Table 1: Quantitative Results of Ascites Segmentation Models. [·] denotes Confidence Interval at 95%. Top half table: performance of models in related works. Bottom half table: proposed method using nnU-Net.

The 2D Residual U-Net was able to achieve an average Dice score of 0.64, with Jaccard, precision, rec./sen., and specificity of 0.50, 0.68, 0.64 and 0.99, respectively. With a CI-95 value in the order of 10^{-2}, the models are fairly robust in their predictions. The performance of the 3D Residual U-Net and 3D BLE-U-Net models, however, were not able to surpass the 2D method. The proposed nnU-Net method achieved the best scores in all of our experiments, attaining an average Dice, Jaccard, precision, rec./sen., and specificity of 0.66, 0.52, 0.67, 0.68 and 0.99, respectively. Compared to the 2D Residual U-Net, it also attained a lower confidence interval indicating a more robust model.

It is desirable, for the application of ascites segmentation, that the model is best trained in 3D. As ascites is a liquid, it does not have a definitive shape prior, and thus, the network learns to segment via texture and intensity. This is evident in Figure 1, which shows a particularly challenging case. The 2D model has segmented parts of the stomach and intestines as ascites, due to ascites having similar intensity profile to water, which is not the case for the 3D models.

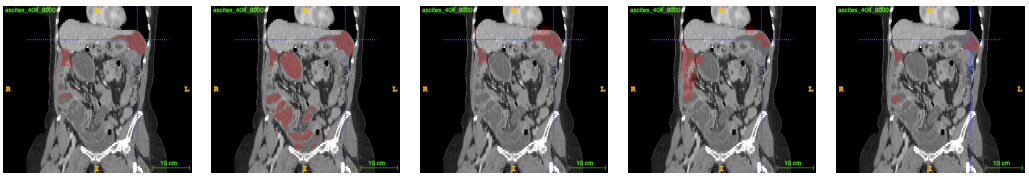

Figure 1: Predicted Ascites on Test Image
{L-to-R: Ground Truth, 2D Residual U-Net, 3D Residual U-Net, 3D BLE-U-Net, nnU-Net}

nnU-Net tends to under-segment regions of ascites, whereas 3D Residual U-Net and BLE-U-Net tends to over-segment. This is evident in the region of ascites in the top-right area of Figure 1, and is reflected by the respective precision score attained. The volume of ascites can be inferred from the predicted segmentation masks, and by definition, an accumulation of more than 25ml of fluid constitutes as a positive case of ascites. The exact volume needn't be precise, but clinical outcomes becomes distinct if thresholded by 1500ml to classify patients into small and large ascites groups [7].

4 Discussion and Conclusion

In this paper, we have presented a 3D segmentation model to accurately segment ascites, in the abdomen and pelvis, for the cancer analysis. As nnU-Net discerns the appropriate training parameters by analyzing the heuristics of the data, it is able to train a network to draw a decision boundary more accurately between ascites and other liquids with similar intensity profile such as water, urine, bile, etc. The network trained was able to achieve an average Dice and precision score of 0.7 and 0.75 respectively. This is higher than state-of-the-art methods in the recent literature.

Acknowledgements

This work was supported by the National Institutes of Health, Clinical Center, National Cancer Institute, and National Institute of Diabetes and Digestive and Kidney Diseases.

Potential Negative Societal Impact Statement

To our knowledge, segmentation of ascites have not been vastly explored in recent literature, as well as the the impact of ascites volume on clinical outcomes. However, existing studies have shown that ascites volume and clinical outcomes can be correlated, thus showing potential of investigation. The alternative to our method would be to use manual segmentation methods such as thresholding, region-growing, etc. But this can be a time consuming and challenging task, as ascites mimics liquid profiles of other bodily fluids around the area. As with any Deep Learning method, it is impossible to guarantee a perfect performing model. We hope the model can provide segmentations as a good starting point, and not to take the prediction by face value.

References

[1] M Chiejina, P Kudaravalli, and H Samant. "Ascites". In: *StatPearls [Internet]* (2022). URL: https://www.ncbi.nlm.nih.gov/books/NBK470482/.

[2] Robert Geirhos et al. "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness". In: *ICLR*. OpenReview.net, 2019.

[3] C Holback et al. "Radiology data from the cancer genome atlas ovarian cancer [tcga-ov] collection". In: *The Cancer Imaging Archive* (2016).

[4] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2 (2021), pp. 203–211.

[5] Hoon Ko et al. "A Deep Residual U-Net Algorithm for Automatic Detection and Quantification of Ascites on Abdominopelvic Computed Tomography Images Acquired in the Emergency Department: Model Development and Validation". In: *Journal of Medical Internet Research* 24.1 (2022), e34415.

[6] Manas K Nag et al. "Body Location Embedded 3D U-Net (BLE-U-Net) for Ovarian Cancer Ascites Segmentation on CT scans". In: *SIPAIM-2022* (2022).

[7] Quan Quan et al. "Relationship between ascites volume and clinical outcomes in epithelial ovarian cancer". In: *Journal of Obstetrics and Gynaecology Research* 47.4 (2021), pp. 1527–1535.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *MICCAI (3)*. Vol. 9351. Lecture Notes in Computer Science. Springer, 2015, pp. 234–241.

[9] J Brian Szender et al. "Impact of ascites volume on clinical outcomes in ovarian cancer: A cohort study". In: *Gynecologic oncology* 146.3 (2017), pp. 491–497.

[10] David J Winkel et al. "Evaluation of an AI-based detection software for acute findings in abdominal computed tomography scans: toward an automated work list prioritization of routine CT examinations". In: *Investigative radiology* 54.1 (2019), pp. 55–59.